RADAR TECNOLÓGICO

Estudio sobre tecnologías emergentes en las industrias de la creación de contenidos digitales para el entretenimiento.

Este estudio ha sido realizado con financiación del Ministerio para la Transformación Digital y de la Función Pública a través de la Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales (SETELECO). El ministerio no comparte necesariamente los contenidos expresados en el mismo, responsabilidad exclusiva de sus autores. Se permite su copia y distribución por cualquier medio siempre que se mantenga el reconocimiento de sus autores, no se haga uso comercial de las obras y no se realice ninguna modificación de las mismas".

GOBIERNO DE ESPAÑA
MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES

#SpainAVSHub

Contenido

Introducción

La Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales (SETELECO) tiene, entre otras competencias, la planificación estratégica y de acción sobre telecomunicaciones, las infraestructuras digitales y los servicios de comunicación audiovisual, así como la elaboración y propuesta de normativa a la ordenación y regulación en estas materias, en consonancia con las disposiciones nacionales, europeas e internacionales vigentes.

La SETELECO se encarga del impulso y refuerzo del sector de producción audiovisual española a través de la internacionalización, el fomento de la innovación y la mejora de la regulación, enmarcado dentro del Componente 25 "España Hub Audiovisual del Europa/Spain AVS Hub" del Plan de Recuperación, Transformación y Resiliencia financiado con los fondos europeos *Next Generation*. Este componente tiene como objetivo "posicionar a España como centro de referencia para la producción audiovisual, mejorando el atractivo de España como plataforma europea de negocio, trabajo e inversión en el ámbito audiovisual, exportador de productos audiovisuales y polo de atracción de talento".

Por otro lado, el Plan "España, Hub Audiovisual de Europa" (Plan "Spain AVS Hub"), presentado en 2021, es uno de los ejes de la agenda España Digital 2025 y tiene como objetivo convertir España en el principal Hub audiovisual de Europa mediante el impulso de la producción audiovisual nacional y la atracción de inversión y actividad económica, el refuerzo de las empresas del sector mejorando su competitividad a través de la digitalización y el apoyo del talento, reduciendo la brecha de género.

La medida 15 de dicho plan, establece el mandato expreso dirigido al actual Ministerio de Transformación Digital y de la Función Pública, de llevar a cabo un seguimiento preciso y adecuado sobre la efectividad de las medidas y, en su caso, la necesidad de reajustes para garantizar la eficacia del conjunto.

Además de conocer la situación del sector se hace necesario en la actualidad tener una visión más amplia no solo de la situación actúa del mismo, sino de su futuro a corto y largo plazo, para lo que es necesario prever las tendencias del mismo.

Es sabido que la tecnología desborda con frecuencia la normativa y obliga a las Administraciones Públicas a una permanente actualización para incorporar nuevos retos o posibilidades no previstas con la tecnología anterior. Es por eso que conocer con anticipación las tendencias tecnológicas puede permitir una mejor preparación de las Administraciones en su papel regulador.

Igualmente se exige cada vez más a las AAPP una dotación de recursos tecnológicos hacia sectores como la educación o la formación que a veces supone adquisiciones de equipamiento con una inminente obsolescencia, ya que las empresas van a preferir a las personas formadas con las últimas tecnologías. Por ello, conocer las tecnologías que pueden hacer obsoletas las actuales da una muy valiosa información a aquellos departamentos con capacidades y responsabilidades en tales dotaciones.

Finalmente, las ayudas y apoyos públicos a los procesos de digitalización en todos los sectores, y con una notable aceleración gracias a los fondos Next Generation, aconseja y hace muy conveniente, conocer qué ayudas pueden ser más estratégicas por apoyar tecnologías que vayan a ser las dominantes en los próximos años.

Por ello se precisa disponer de un análisis de carácter tecnológico de dichas tendencias, de las tendencias de la tecnología en los subsectores audiovisuales más dinámicos, su grado de madurez y aplicabilidad en este ejercicio. Este informe se denominará Radar tecnológico.

La industria de generación de contenidos digitales para el entretenimiento se caracteriza por una actitud favorable a la adopción de nuevas tecnologías y la fuerte influencia que estas tienen a la hora de generar brechas importantes en la competitividad de las empresas.

Por otro lado, debido a la aceleración en la innovación tecnológica, muchas empresas y entidades públicas tienen dificultades para mantenerse al día y evaluar con agilidad cuándo y cómo introducir en sus procesos de producción nuevas técnicas o herramientas. Esta situación maximiza el valor que un radar tecnológico puede aportar al sector.

Un radar tecnológico se caracteriza por resumir una gran cantidad de información en un solo diagrama y permitir a las Administraciones Públicas conocer la evolución del sector para graduar las políticas de apoyo a desarrollar y a las empresas del sector definir y evaluar su estrategia tecnológica a corto y largo plazo.

Entre otros, podemos identificar los siguientes casos de uso de esta herramienta:

- Auditar la cartera de tecnología de una empresa o entidad, permitiendo identificar riesgos, amenazas y oportunidades derivadas del uso de la tecnología en sus operaciones.
- Ayuda a tomar decisiones a la hora de invertir o abandonar tecnologías que están en desuso, facilitando y acelerando la adopción de las tecnologías que están dando mejores resultados en la industria.
- Mantenerse al día de las nuevas tecnologías en innovaciones que tiene potencial ayudando desde las administraciones públicas a las empresas del sector a posicionarse con rapidez y obtener ventajas competitivas.

En este sentido, la Administración, consciente de la necesidad de la constante adopción de nuevas tecnologías y desarrollos tecnológicos por parte del sector de la creación audiovisual, considera necesaria la creación de este radar tecnológico para poder profundizar en el conocimiento de la tecnología en cuestión de identificar documentación, herramientas comerciales o de código abierto que faciliten su evaluación, prueba o adopción por parte del sector en sus procesos de creación digital.

Planteamiento

La industria de la animación, los efectos visuales (VFX) y videojuegos constituyen un ámbito altamente competitivo en el que, el talento y la tecnología son elementos esenciales para obtener una ventaja sobre la competencia. Los gastos relacionados con el personal, el software y el hardware representan los principales costos que impactan en los beneficios empresariales, requiriendo una gestión meticulosa y precisa.

Este documento analiza el papel crucial de la tecnología y cómo una estrategia bien estructurada y revisada periódicamente, puede influir significativamente en la creación de nuevos servicios y en la reducción de los costos de producción de contenidos existentes.

Importancia de la Tecnología y la Estrategia

El sector destaca por su capacidad para adoptar rápidamente nuevas tecnologías, un rasgo distintivo de la industria. Sin embargo, la falta de un marco definido para la innovación puede llevar a decisiones de inversión en herramientas de hardware y software basadas en tendencias y en información poco fiable.

Es vital que tanto las grandes como las pequeñas empresas dediquen tiempo a evaluar sus inversiones tecnológicas. Este proceso permite identificar nuevas herramientas y metodologías que merecen inversión y determinar cuáles de las actualmente utilizadas deben ser descartadas por no aportar valor añadido.

Este análisis es complejo y requiere una inversión considerable de tiempo, conocimiento y recursos. La falta de marcos de trabajo y de fuentes de información actualizadas y fiables añade dificultades adicionales. En respuesta a estos desafíos, ThoughtWorks publicó en 2010 el primer radar tecnológico, una herramienta diseñada para identificar tecnologías emergentes relevantes, revelar riesgos y oportunidades, y gestionar una estrategia tecnológica efectiva.

Uno de los objetivos de este documento es utilizar el radar tecnológico de *ThoughtWorks* para facilitar a las empresas del sector en España, la elaboración de una estrategia de innovación y la toma de decisiones respecto a la adopción de nuevas metodologías y herramientas de software y hardware.

Irrupción vertiginosa de nuevas herramientas tecnológicas

En los últimos dos años, la irrupción de tecnologías innovadoras, como la inteligencia artificial o el big data entre otras muchas, ha transformado dramáticamente el ecosistema tecnológico de todas las industrias. Incluyendo, como no, al sector de la creación de contenidos digitales.

Este nuevo escenario tremendamente cambiante genera incertidumbre, ansiedad y oportunidades, especialmente en la industria de la creación de contenido digital para el entretenimiento. Estas áreas tecnológicas han evolucionado a una velocidad exponencial en los últimos años, factores que dificultan a las pequeñas y medianas empresas del sector en España acceder a la información necesaria, en el formato adecuado, para decidir dónde invertir sus esfuerzos y recursos, en muchos casos limitados.

El objetivo de esta iniciativa es identificar aplicaciones prácticas donde estas nuevas tecnologías estén a punto de alcanzar su madurez y presentar estos casos en un formato de radar tecnológico que facilite su difusión, comprensión y adopción.

Objetivos y Metodología

Este documento está dirigido principalmente a ejecutivos y directores de tecnología de empresas españolas del sector que se plantean las siguientes cuestiones a la hora de definir su estrategia de inversión tecnológica del próximo año:

1. ¿En qué flujos de trabajo aportar valor y reducir costos?
2. ¿En qué soluciones invertir?
3. ¿Cuáles son los primeros pasos para llevar a cabo dicha inversión?

Este documento tiene como objetivo proporcionar a las administraciones públicas un conocimiento de las tendencias tecnológicas del sector para abordar con solvencia tres responsabilidades que le incumben:

- Regulatorias
- Dotación de recursos
- Apoyos públicos

Para facilitar la toma de decisiones informada sobre estas cuestiones, este documento tiene como objetivo la creación de un radar tecnológico que representará de forma gráfica las tecnologías que pueden impactar la estrategia de los estudios de creación de contenidos digitales en los próximos años.

Las tecnologías incluidas en el radar se agruparán en 4 categorías representadas por cuatro cuadrantes:

- **Técnicas**: Procesos y metodologías de producción de contenidos digitales audiovisuales
- **Herramientas**: Software y hardware que usan los artistas y técnicos de la industria en su día a día tales como: soluciones de modelado y animación 3D, editores y compositores de imágenes o tabletas, monitores y *lidars*.
- **Lenguajes y *frameworks***: Lenguajes de programación y scripting como C++ o Python y librerías y *frameworks* usados para producir herramientas.

- **Plataformas e Infraestructura**: Tecnologías hardware y software sobre las que se despliegan las herramientas de producción de contenidos audiovisuales como portátiles, estaciones de trabajo, sistemas de almacenamiento, sistemas operativos, etc.

Por cada una de las tecnologías posicionadas en el radar, se incluirán los siguientes apartados:

- Descripción de la tecnología en el contexto de aplicación de la industria.
- Casos de uso de éxito y áreas de mejora.
- Información sobre la titularidad de las tecnologías, las condiciones de uso y privacidad.
- Información sobre su disponibilidad y costes de Implantación.

Por último, cada tecnología se posicionará en un anillo describiendo su grado de madurez para ser utilizada en proyectos de producción. El radar contendrá 4 anillos:

- **Adoptar**: La tecnología está preparada para ser adoptada por la industria.
- **Probar**: La tecnología está preparada para probarse en proyectos de bajo riesgo.
- **Evaluar**: La tecnología debería comenzar a ser investigada para identificar áreas de la producción en las que puedan aportar valor.
- **Evitar**: La tecnología no está preparada para tomar posiciones en ella.

Una vez completado el radar tecnológico permitirá a sus usuarios y lectores:

- Reducir el esfuerzo necesario para tomar decisiones importantes sobre el panorama tecnológico de la industria.
- Evitar el riesgo de operar dentro de una burbuja desaprovechando las oportunidades que nuevas tecnologías pueden aportar a la industria
- Analizar una gran cantidad de tecnologías de manera más efectiva e identificar de forma temprana las tecnologías más relevantes para su industria.
- Informar sobre tecnologías relevantes de una manera que todas las personas implicadas en la definición de la estrategia tecnológica puedan entender y respaldar

Para identificar las tecnologías con mayor potencial de impacto en la industria y posicionarlas en el radar, se formará un panel de profesionales con amplia experiencia en el sector y en la aplicación de soluciones de IA en la industria española. Este panel será responsable de identificar los casos de uso con mayor potencial y desarrollar en detalle las recomendaciones para cada uno de ellos.

Es fundamental que la información incluida en este documento se base en análisis y experiencias de aplicación en proyectos cercanos a la realidad del negocio. Asimismo, es crucial que estos proyectos se hayan llevado a cabo en el contexto español y europeo, ya que en el ámbito de la IA este marco tiene un gran impacto en la viabilidad de las

soluciones. Por esta razón, ambos criterios serán utilizados para priorizar los casos de uso.

Autores

El *Technology Advisory Board*, responsable de la redacción de este informe está formado por un grupo de profesionales con una contrastada experiencia y que, a través de su visión y experiencia, analizan y plasman en este documento algunas de las tendencias tecnológicas más vanguardistas actualmente.

Debido al impacto transversal de estas tecnologías, todos los autores de este documento han participado en la elaboración de los diversos apartados que forman parte de este radar tecnológico.

- Manuel Meijide. Director - Mundos Digitales
- Alberto Arenas. *Principal Pipeline Engineer* – Hogarth
- Víctor M. Feliz. *Founder* – Motiva CG
- Alba Meijide. *AI Consultant & Project Manager*
- Adrián Pueyo. *Head of VFX & VP* – Orca Studios

Radar Tecnológico

Técnicas

Rotoscopia Automatizada

Introducción

La tarea de rotoscopia consiste en crear contornos precisos de elementos que permiten recortar y separar un elemento de una imagen. Es una de las tareas más importantes en todo el proceso de generación de efectos visuales y también es una de las más costosas en tiempo y recursos. La integración de sistemas de inteligencia artificial (IA) en las tareas de rotoscopia promete ser una gran revolución en el ámbito de los efectos visuales, animación y videojuegos. Estos avances tecnológicos permiten automatizar procesos tradicionalmente manuales y laboriosos, logrando una precisión y eficiencia sin precedentes.

En este contexto, los sistemas de IA para segmentación juegan un papel crucial al asistir en esta tarea. Estos sistemas se encargan de dividir imágenes o datos en partes significativas y más fáciles de analizar, facilitando la identificación y clasificación de objetos específicos dentro de una imagen o vídeo. Los orígenes de la segmentación se remontan a los inicios de la visión por computadora y el procesamiento de imágenes digitales en las décadas de 1960 y 1970, cuando los investigadores comenzaron a desarrollar algoritmos básicos para detectar bordes y áreas homogéneas en imágenes.

A lo largo de los años, esta área ha evolucionado significativamente con la incorporación de técnicas avanzadas de aprendizaje automático y redes neuronales profundas. Estas innovaciones han mejorado drásticamente la precisión y eficiencia de la segmentación, haciendo posibles aplicaciones complejas como la rotoscopia automatizada. Sin embargo, la segmentación sigue siendo un gran reto debido a la variabilidad y complejidad inherente de las imágenes del mundo real. Factores como la iluminación, el ruido, las sombras y las diferencias en la escala y perspectiva de los objetos complican la tarea de dividir las imágenes de manera precisa y consistente.

A pesar de los avances, lograr una segmentación precisa en contextos complejos y en tiempo real sigue siendo un objetivo desafiante y en constante evolución en el campo de la inteligencia artificial. No obstante, los progresos continuos en esta área auguran un futuro donde la rotoscopia, apoyada por sistemas de IA avanzados, transformará radicalmente los procesos de creación de efectos visuales, animación y videojuegos, reduciendo significativamente los costos y tiempos asociados, y permitiendo una mayor creatividad y calidad en las producciones audiovisuales.

Hasta ahora, las soluciones existentes para la rotoscopia asistida por IA han sido predominantemente de código abierto, ofreciendo herramientas que permiten a los desarrolladores y artistas experimentar con algoritmos de segmentación y rotoscopia. Estas soluciones han proporcionado una base sólida para la investigación y el desarrollo en el campo. Sin embargo, recientemente hemos comenzado a ver la aparición de

tecnologías y softwares comerciales específicamente diseñados para la rotoscopia asistida por IA, dirigidos al sector profesional. Estas nuevas herramientas prometen integrar de manera más fluida en los flujos de trabajo de producción, ofreciendo una mayor facilidad de uso y resultados más consistentes y precisos, lo que marca un paso significativo hacia la adopción generalizada de la IA en la industria de los efectos visuales.

Descripción

Tradicionalmente la tarea de rotoscopia se lleva a cabo definiendo el contorno del elemento a recortar punto a punto generando una curva bezier. Dependiendo de la complejidad del contorno y de la precisión necesaria para definirlo, esto puede requerir más de 20 o 30 puntos que deben ajustarse manualmente uno a uno

Con el uso de Redes Neuronales de segmentación pre-entrenadas el proceso de generación de las máscaras sólo necesita de la selección de 4 o 5 puntos de inclusión que sirvan de muestra del color o textura característicos del elemento y de puntos de exclusión si existen elementos que se superponen a elemento y que no son parte de él. Como se muestra en las siguientes figuras la reducción del coste de generación de las siluetas es muy significativo:

Fig 1. A la izquierda se muestra la silueta generada con más de 30 puntos posicionados de forma manual cuyas tangentes han tenido que ser editadas manualmente. A la derecha se muestra el resultado de la silueta generada por una red neuronal de segmentación a partir de 3 puntos de inclusión y 2 de exclusión

Las herramientas de rotoscopia asistida por IA están diseñadas para mejorar significativamente la eficiencia y precisión del proceso de rotoscopia en la producción de efectos visuales, animación y videojuegos. A continuación, se describen tres funcionalidades principales de estas herramientas y su relación con la tarea de rotoscopia.

1. Segmentación Automática de Objetos

La segmentación automática de objetos es una funcionalidad clave en las herramientas de rotoscopia asistida por IA. Utilizando algoritmos avanzados de aprendizaje automático y redes neuronales, la herramienta puede identificar y delinear automáticamente los contornos de los objetos en cada fotograma de una secuencia de video. Esta capacidad reduce drásticamente el tiempo necesario para crear máscaras precisas, ya que elimina la necesidad de trazar manualmente los bordes de los objetos. En la tarea de rotoscopia, esto significa que los artistas pueden generar contornos precisos de elementos en movimiento de manera rápida y eficiente, permitiendo una integración más fluida de efectos visuales.

2. Rastreo de Objetos en Movimiento

El rastreo de objetos en movimiento es otra funcionalidad esencial. Esta característica permite a la herramienta seguir automáticamente el movimiento de los objetos a lo largo de varios fotogramas, ajustando dinámicamente las máscaras para mantener la precisión del contorno. Al aprovechar técnicas de seguimiento basadas en IA, esta funcionalidad minimiza el esfuerzo manual requerido para ajustar las máscaras en cada fotograma individualmente. Para la rotoscopia, esto significa que los artistas pueden asegurar una coherencia y precisión en la separación de elementos en movimiento, incluso en escenas complejas con cambios rápidos y movimientos irregulares.

3. Corrección y Refinamiento Interactivo

A pesar de los avances en la automatización, la intervención manual sigue siendo necesaria para garantizar la máxima precisión. Las herramientas de rotoscopia asistida por IA ofrecen funcionalidades de corrección y refinamiento interactivo, donde los artistas pueden ajustar manualmente las máscaras generadas automáticamente. Esta característica incluye herramientas de edición intuitivas que permiten modificar contornos, ajustar áreas específicas y corregir errores detectados por la IA. Esta funcionalidad es crucial para afinar los detalles y asegurar que los resultados finales cumplan con los estándares de calidad exigidos en la producción profesional de efectos visuales.

Casos de Uso

La rotoscopia asistida por IA tiene aplicaciones prácticas diversas y específicas que destacan su potencial para transformar la producción audiovisual. En esta sección, se explorarán los casos de uso ideales donde esta tecnología proporciona un valor significativo, así como las limitaciones y áreas de mejora donde la tecnología aún enfrenta desafíos. Esta evaluación integral ayudará a entender mejor las capacidades actuales y futuras de la rotoscopia asistida por IA.

Casos de Uso Ideales:

1. Objetos que se Distinguen Claramente del Fondo

   La tecnología de rotoscopia asistida por IA es extremadamente efectiva cuando los objetos a segmentar tienen colores y texturas que contrastan significativamente con el fondo. En estos casos los sistemas asistidos por IA permiten una segmentación rápida y precisa, reduciendo el tiempo necesario para la rotoscopia manual.

2. Escenas con Iluminación Consistente

   Las escenas con una iluminación uniforme y predecible son ideales para la segmentación automatizada, ya que los algoritmos de IA pueden detectar bordes y contornos sin interferencias de sombras o variaciones de luz. Estas escenas mejoran la precisión de la segmentación, minimizando la necesidad de correcciones manuales.

3. Tomas Estáticas o con Movimiento Suave

   En tomas donde el movimiento de la cámara y de los objetos es lento y predecible, los sistemas de IA pueden rastrear y segmentar con gran precisión, facilitando el seguimiento continuo de los objetos, y asegurando que las máscaras sean coherentes a lo largo de la secuencia.

4. Personajes o Objetos con Bordes Definidos

   Los sistemas de IA para segmentación son particularmente eficaces a la hora de segmentar personajes o objetos con contornos nítidos y bien definidos.

Limitaciones

1. Escenas con Iluminación Compleja

   Los sistemas de segmentación automatizada pueden experimentar dificultades en escenas con iluminación variable, sombras pronunciadas o reflejos. Estos casos destacan la necesidad de algoritmos más robustos que puedan manejar condiciones de iluminación difíciles sin pérdida de precisión.

2. Objetos con Bordes Difusos o Transparentes

   Los objetos con bordes difusos, transparencias o materiales translúcidos son un desafío para la segmentación automatizada. Este tipo de objetos presentan contornos poco definidos y varían en opacidad, lo que dificulta a los algoritmos de IA determinar con precisión dónde termina el objeto y comienza el fondo. Además, los materiales translúcidos pueden causar distorsiones visuales y reflejos complejos, complicando aún más el proceso de segmentación.

3. Escenas con Lluvia, Nieve u Oclusiones

   La presencia de elementos como lluvia, nieve o cualquier objeto que ocluya parcialmente los elementos de interés puede complicar la segmentación. Para que estos sistemas de IA puedan segmentar con éxito estas escenas son necesarias mejoras en el manejo de entornos dinámicos y oclusiones.

4. Movimiento Rápido o Errático

   En escenas donde hay movimiento muy rápido o errático, la IA puede tener dificultades para rastrear y segmentar objetos con precisión. Para gestionar estas escenas, las mejoras deben aplicarse en los algoritmos de seguimiento, no de segmentación, para que estos puedan gestionar movimientos complejos y rápidos sin errores en el seguimiento.

5. Dependencia de Corrección Manual

   Aunque la tecnología automatiza gran parte del proceso, aún requiere intervención manual para ajustes finos y corrección de errores. La segmentación automática puede producir resultados que no son perfectos, especialmente en escenas complejas con múltiples elementos, iluminación variable o movimientos rápidos. Los algoritmos de IA, aunque avanzados, todavía pueden cometer errores, como no captar detalles minuciosos, fallar en rastrear objetos de manera consistente a lo largo de varios fotogramas, o manejar de manera inadecuada bordes complejos y transparencias. Por lo tanto, los artistas deben revisar y ajustar manualmente las máscaras generadas, refinando contornos, corrigiendo áreas mal segmentadas y asegurando que todos los detalles relevantes sean precisos. Esta intervención manual es crucial para mantener la alta calidad requerida en producciones profesionales, lo que implica que, a pesar de los avances en la automatización, la habilidad y la experiencia humana siguen siendo esenciales para obtener resultados óptimos.

Titularidad o Patentes

Cada sistema de rotoscopia asistida por IA tiene sus propios métodos de gestión de titularidad y patentes, dependiendo de si se trata de software de código abierto, software comercial u otras modalidades.

Los sistemas de rotoscopia asistida por IA de **código abierto** suelen estar disponibles para su uso y modificación bajo licencias específicas que definen los términos de uso, distribución y modificación. Estas licencias pueden variar, pero generalmente permiten a los usuarios adaptar y mejorar el software, contribuyendo a una comunidad más amplia de desarrolladores y usuarios.

Por otro lado, los sistemas de rotoscopia asistida por IA **comerciales** están protegidos por patentes y derechos de propiedad intelectual que pertenecen a las empresas que los desarrollan. Estos sistemas suelen requerir licencias pagadas para su uso y están sujetos a términos estrictos de uso que protegen la propiedad intelectual del desarrollador. Las patentes pueden cubrir una amplia gama de aspectos, desde algoritmos específicos utilizados para la segmentación y el seguimiento hasta interfaces de usuario y métodos de integración en flujos de trabajo de producción.

En este documento, se incluyen referencias a herramientas específicas de rotoscopia asistida por IA. Para cada una de estas herramientas, se considerarán las particularidades respecto a la titularidad y las patentes, proporcionando información sobre sus licencias, términos de uso y cualquier patente relevante que pueda afectar su implementación y utilización. Esta información es crucial para entender las implicaciones legales y de uso asociadas con cada herramienta, permitiendo a los usuarios tomar decisiones informadas sobre qué sistema de rotoscopia asistida por IA es más adecuado para sus necesidades específicas.

Disponibilidad y Costes de Implantación

En esta sección se presentan las distintas opciones para adoptar soluciones de inteligencia artificial (IA) para la automatización parcial de la tarea de rotoscopia, cada una de ellas con sus ventajas y sus limitaciones. Con el avance de la IA, se han desarrollado herramientas que facilitan y aceleran este proceso, sin embargo, la elección de la solución adecuada debe hacerse con cuidadosa consideración.

Es crucial adoptar la solución que más se adapte a las necesidades específicas de cada proyecto y empresa. Esto no solo implica evaluar:

- La precisión de las máscaras generadas, que es un factor fundamental para la calidad del producto final
- La capacidad técnica de los recursos disponibles para implementar esta innovación.
- Los recursos computacionales necesarios para procesar grandes volúmenes de datos

- La seguridad de la información procesada son aspectos que no deben subestimarse.
- Costos y la capacidad técnica del equipo encargado.

Existen tres opciones principales a través de las cuales podemos incorporar herramientas de IA para asistir las tareas de rotoscopia:

- Librerías y Modelos Open-source
  Los modelos open-source se refieren a aquellos algoritmos y herramientas de inteligencia artificial cuyo código fuente es accesible al público, permitiendo a los usuarios estudiar, modificar y distribuir el software según sus necesidades. En el contexto de la rotoscopia, los modelos open-source ofrecen una base sólida para el desarrollo de soluciones personalizadas debido a su flexibilidad y la posibilidad de adaptarlos a requisitos específicos. Estos modelos y librerías incluyen: Segment Anything, YOLOv8 and OpenCV

- Herramientas comerciales alojadas en la nube.
  En el ámbito de la rotoscopia y segmentación, existen diversas herramientas comercializadas que operan en línea, facilitando el proceso mediante soluciones basadas en la nube. Estas herramientas están diseñadas para llevar a cabo tareas específicas de segmentación y ofrecen una alternativa accesible para estudios y profesionales que buscan implementar tecnologías de inteligencia artificial sin incurrir en los costos y la complejidad de desarrollar soluciones personalizadas. Estás herramientas incluyen: RunwayML, Sunscreen y Mask Prompter.

- Herramientas comerciales a media ejecutadas en infraestructuras propias**.**
  El desarrollo de herramientas ad-hoc específicamente diseñadas para tareas de rotoscopia en el sector audiovisual representa una solución avanzada y precisa para satisfacer las necesidades exigentes de la industria de la animación y los efectos visuales. Estas herramientas específicas, y por el momento escasas, deben estar diseñadas para ofrecer un rendimiento superior y una exactitud extrema en la creación de máscaras, superando las limitaciones de los modelos genéricos y comerciales. Un ejemplo destacado de estas herramientas es SequoIA

La disponibilidad y los costes de implantación de los sistemas de rotoscopia asistida por IA varían considerablemente dependiendo de cada herramienta y su modelo de distribución. En las siguientes secciones del documento, se detalla la disponibilidad y los costes específicos de implantación para cada herramienta, proporcionando una visión clara y detallada que permitirá evaluar las opciones más adecuadas según sus necesidades y presupuestos.

Motion Tracking

Introducción

El *Motion Tracking* es una tecnología avanzada diseñada para capturar y seguir movimientos dentro de secuencias de vídeo con alta precisión. Esta tecnología utiliza algoritmos de inteligencia artificial para identificar y rastrear objetos en movimiento, facilitando su integración en una variedad de aplicaciones como efectos visuales, animación, videojuegos y sistemas de vigilancia. La tecnología comenzó a desarrollarse de manera significativa a principios de la década de 2010, impulsada por los avances en el aprendizaje automático y la visión por computadora. El *Motion Tracking* asistido por IA empezó a ganar tracción en la industria alrededor de 2012, coincidiendo con el auge de las redes neuronales profundas y el aprendizaje profundo.

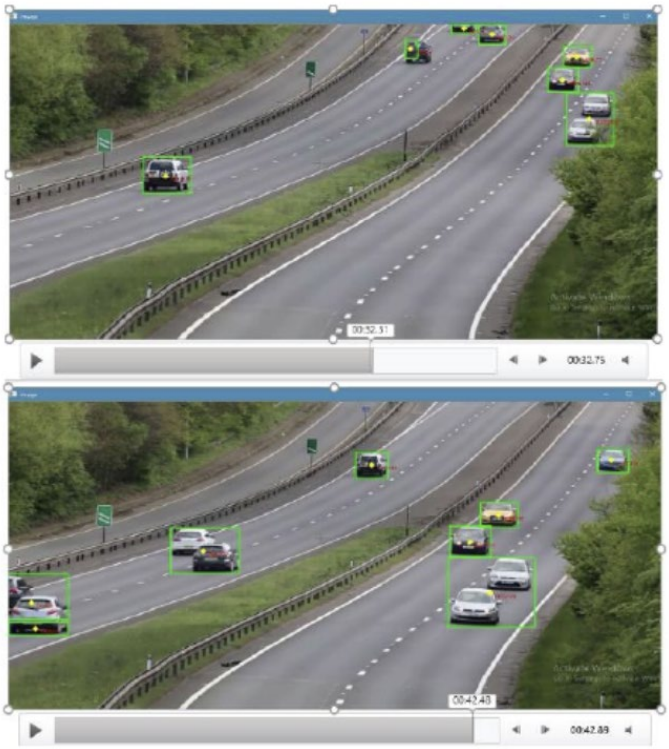


Fig 2. Ejemplo de motion tracking con algoritmos de IA aplicada a la vigilancia y seguridad, presentada en el artículo de investigación oficial "Object Detection and Tracking using Deep Learning and Artificial Intelligence for Video Surveillance Applications" [14]

Esta tecnología se relaciona estrechamente con la tarea de rotoscopia y los sistemas de IA para segmentación, ya que ambos procesos implican la identificación y el seguimiento de objetos en movimiento dentro de secuencias de vídeo. Mientras que la rotoscopia se centra en crear contornos precisos de elementos específicos para su posterior manipulación o eliminación del fondo, el *motion tracking* utiliza algoritmos avanzados para seguir esos elementos a lo largo de múltiples fotogramas, asegurando

que las máscaras y los contornos se mantienen coherentes incluso cuando los objetos se mueven rápidamente o cambian de forma. Esta sinergia entre *motion tracking* y rotoscopia asistida por IA permite una mayor eficiencia y precisión en la producción de efectos visuales, facilitando tareas tradicionalmente laboriosas y propensas a errores.

Descripción

El *Motion Tracking* asistido por IA se integra eficazmente con las tareas de rotoscopia y segmentación, optimizando y automatizando procesos que tradicionalmente han sido laboriosos y exigentes. A continuación, se presentan tres funcionalidades principales de esta herramienta y su aplicación en la rotoscopia y segmentación por IA:

1. Seguimiento Automático de Objetos

Los sistemas de IA para tracking utilizan algoritmos avanzados para identificar y rastrear objetos en movimiento a lo largo de múltiples fotogramas. Esto permite asegurar que las máscaras y contornos generados automáticamente se mantienen coherentes y precisos a lo largo de toda la secuencia de video, reduciendo la necesidad de ajustes manuales. Esto facilita la segmentación dinámica de objetos, adaptándose a cambios de posición, forma y velocidad de los elementos rastreados.

2. Corrección Automática de Errores

Estos algoritmos Implementan técnicas de aprendizaje automático para detectar y corregir automáticamente errores comunes en la rotoscopia y segmentación, como desalineaciones o fragmentaciones de la máscara. El objetivo es minimizar la intervención manual al corregir automáticamente las máscaras y contornos, manteniendo una alta precisión y coherencia en escenas complejas.

3. Integración con Herramientas de Edición

Los desarrollos de IA para seguimiento o *tracking*, se integra perfectamente con las principales plataformas de edición de video y efectos visuales, proporcionando un flujo de trabajo optimizado y herramientas intuitivas para la rotoscopia y segmentación. El desarrollo y avance de estos algoritmos proporciona herramientas interactivas para ajustar y perfeccionar las segmentaciones automáticas, permitiendo un control preciso sobre los resultados finales y mejorando la calidad general de la producción.

Estas funcionalidades principales del *Motion Tracking* asistido por IA, aplicadas a la rotoscopia y segmentación, no solo mejoran la eficiencia y precisión de estos procesos, sino que también liberan a los artistas de tareas tediosas, permitiéndoles enfocarse en aspectos más creativos y de valor añadido en sus proyectos.

Casos de Uso

El *Motion Tracking* asistido por IA, al igual que las soluciones de IA para segmentación, ofrece una gama de aplicaciones donde puede mejorar significativamente la eficiencia y precisión de las tareas de seguimiento y segmentación de objetos en video. A continuación, se presentan algunos casos de uso ideales y limitaciones específicas de la tecnología. Es importante destacar que muchos de estos casos de uso y limitaciones son comunes a ambas tecnologías, reflejando la interdependencia y sinergia entre el *motion tracking* y la segmentación asistida por IA en la producción de efectos visuales y animación.

Casos de Uso Ideales

1. Movimientos Uniformes de Cámara

   La tecnología de *motion tracking* asistido por IA es altamente efectiva en escenas donde la cámara se mueve de manera suave y predecible, como paneos, *travellings* o tomas con *steadycam*. Estos sistemas garantizan un seguimiento preciso y coherente de los objetos en movimiento, facilitando la rotoscopia y segmentación sin necesidad de ajustes manuales extensivos.

2. Objetos que Contrastan con el Fondo

   Estas tecnologías de seguimiento de objetos funcionan excelentemente cuando los objetos a rastrear tienen un color o textura que contrasta claramente con el fondo, lo que facilita la detección y seguimiento por los algoritmos de IA, aumentando la precisión y velocidad del *tracking*, y reduciendo la necesidad de intervención manual y mejorando la calidad de la segmentación.

3. Escenas con Iluminación Consistente

   Esta tecnología es ideal para escenarios con iluminación estable y uniforme, donde no hay cambios drásticos en la luz o sombras que puedan confundir los algoritmos de seguimiento. Estas condiciones permiten mantenerla precisión del *tracking* a lo largo de toda la secuencia, asegurando que los objetos se mantengan correctamente segmentados y rastreados.

Limitaciones

1. Escenas con Iluminación Variable

   Los desarrollos de seguimiento a través de IA pueden tener dificultades en escenarios con iluminación cambiante, sombras intensas o reflejos que confunden los algoritmos de seguimiento. Estos cambios pueden alterar la apariencia de los objetos, haciendo que los algoritmos pierdan precisión en la

identificación y rastreo de los contornos. La variabilidad en la luz puede crear sombras dinámicas y reflejos no previstos, que los modelos de IA actuales no siempre pueden manejar eficazmente. Esto puede resultar en errores de seguimiento, donde los objetos son mal identificados o se pierden temporalmente, requiriendo una intervención manual significativa para corregir estos problemas y garantizar la continuidad del seguimiento.

2. Objetos con Bordes Difusos o Transparentes

Los objetos con bordes suaves, difusos o transparencias presentan un desafío significativo para la tecnología de tracking. Esto se debe a que los algoritmos de IA pueden tener dificultades para identificar y seguir con precisión los contornos de estos objetos, ya que los bordes poco definidos y las áreas translúcidas pueden confundirse con el fondo o con otros objetos cercanos. Además, los materiales transparentes pueden crear distorsiones visuales y reflejos que complican aún más la tarea de seguimiento, lo que a menudo resulta en una segmentación imprecisa y en la necesidad

3. Movimientos Rápidos e Irregulares

Escenas con movimientos rápidos, erráticos o impredecibles pueden ser problemáticas, ya que los algoritmos pueden no rastrear adecuadamente. Estos movimientos abruptos y veloces pueden causar que el algoritmo pierda el seguimiento del objeto, generando errores en la segmentación y en la continuidad del tracking. Además, los cambios repentinos en la dirección o velocidad del objeto pueden superar la capacidad de respuesta de los algoritmos actuales, resultando en saltos o inconsistencias en el seguimiento que requieren intervención manual para corregirse. Este tipo de escenas desafían los límites de la tecnología actual y resaltan la necesidad de desarrollar algoritmos más robustos y adaptativos que puedan manejar movimientos complejos y no predecibles con mayor precisión.

4. Ambientes Complejos y Texturas Similares

En ambientes con texturas muy similares o complejas, los algoritmos de IA pueden tener dificultades para distinguir y seguir los objetos correctamente. Esta situación ocurre frecuentemente en escenarios donde el fondo y los objetos en movimiento tienen patrones o colores que se asemejan mucho, lo que confunde a los algoritmos de seguimiento. La falta de contrastes claros entre los elementos puede llevar a errores en la detección y seguimiento, causando que los objetos se pierdan o que el tracking se vuelva impreciso. Este desafío es especialmente relevante en entornos naturales, como bosques densos o paisajes urbanos con mucha repetición de elementos visuales, donde la tecnología necesita ser más sofisticada para diferenciar los objetos de interés del fondo.

Titularidad o Patentes

La titularidad y las patentes de las herramientas de IA para *motion tracking* varían según el caso de uso y la herramienta específica. Cada solución de *motion tracking* asistido por IA puede tener diferentes métodos de gestión de propiedad intelectual, dependiendo de si se trata de software de código abierto, *software* comercial o soluciones patentadas por empresas particulares. Las herramientas de código abierto suelen estar disponibles bajo licencias específicas que permiten su uso y modificación, mientras que las soluciones comerciales están protegidas por patentes y derechos de propiedad intelectual que pertenecen a las empresas desarrolladoras. En el documento se incluye una herramienta de código abierto para este propósito, RLOF, la cual ofrece flexibilidad y accesibilidad a los usuarios interesados en soluciones libres y modulares.

Disponibilidad y Costes de Implantación

La disponibilidad y los costes de implantación de las herramientas de IA para *motion tracking* varían considerablemente según el caso de uso específico y la herramienta seleccionada. Factores como si la solución es de código abierto o comercial, la complejidad del proyecto, la infraestructura tecnológica existente, y las necesidades específicas del usuario influyen en los costes y la accesibilidad. Por ejemplo, las herramientas de código abierto pueden ser más accesibles desde el punto de vista económico, ya que generalmente no requieren un pago inicial por licencias. Sin embargo, pueden requerir una mayor inversión en tiempo y recursos para su personalización y mantenimiento, ya que su implementación puede necesitar conocimientos técnicos especializados y esfuerzos continuos para optimizar y adaptar la herramienta a necesidades particulares. En este documento, se explicarán los costes asociados a una de las herramientas de código abierto para *motion tracking*, proporcionando un análisis detallado que permitirá evaluar su viabilidad en comparación con soluciones comerciales.

Captura de entornos con *Radiance Fields*

Introducción

En este apartado hablaremos de dos técnicas que si bien tienen un mismo objetivo y parten de un mismo conjunto de datos son claramente diferenciables: *NeRFs* y *Gaussian Splatting* (a menudo también llamado *Gaussian Splatter*) ambas tienen en común ser alternativas a la fotogrametría a la hora de generar modelos tridimensionales a partir de imágenes, donde se sustituyen las técnicas habituales de representación de los espacios mediante vértices, polígonos y texturas por un enfoque radicalmente distinto.

La fotogrametría tradicional, aunque efectiva y a día de hoy insuperable en algunos aspectos, se basa en la reconstrucción de modelos 3D creando una malla formada por vértices y caras (generalmente triángulos) que luego se texturiza para

simular la apariencia de la escena original, mientras que tanto *NeRFs* como *Gaussian Splatting* proponen técnicas alternativas para representar estos espacios.

Descripción

Neural Radiance Fields (*NeRFs*) se presentó por primera vez en 2020 en el artículo de investigación oficial, *"NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis"* si bien ya anteriormente se habían tratado de conseguir resultados similares mediante el uso de redes neuronales pero con grandes limitaciones tanto de tiempos de entrenamiento e inferencia como de uso de memoria.

El enfoque de los *NeRFs* se basa en el uso de redes neuronales para aprender una función de representación continua de la escena tridimensional a partir de un conjunto de imágenes 2D tomadas desde diferentes ángulos de la escena. Una vez entrenada, la red puede sintetizar nuevas vistas de la escena desde cualquier perspectiva, incluyendo efectos complejos como los reflejos, refracciones o luces caústicas.

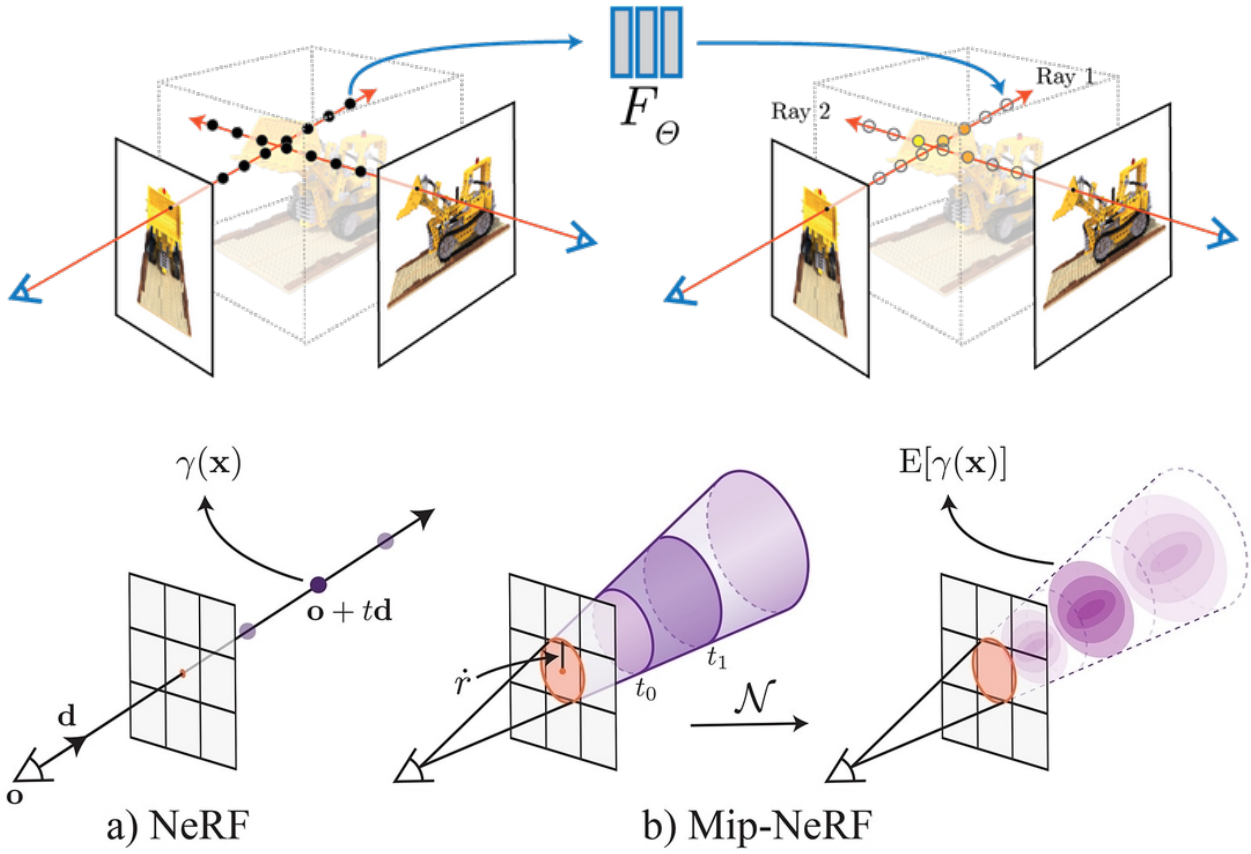


Fig 3. Ilustración funcionamiento NeRF

Gaussian Splatting se basa en las técnicas de *splatting* de las que ya se puede encontrar bibliografía en las décadas de 1980 y 1990 como métodos para la representación tanto de imágenes como de superficies como, por ejemplo: *"A Fast*

Algorithm for Volume Rendering Using a Shear-Warp Factorization of the Viewing Transformation" en el año 1994 para usos médicos. Pero para usos prácticos siempre nos referiremos al uso dado en el artículo de investigación de 2023 *"3D Gaussian Splatting for Real-Time Radiance Field Rendering"* donde se genera una nube de *Splats* que representan la radiancia de ese punto encapsulando unas funciones que se asemejan a técnicas bien conocidas en los videojuegos como los harmónicos esféricos.

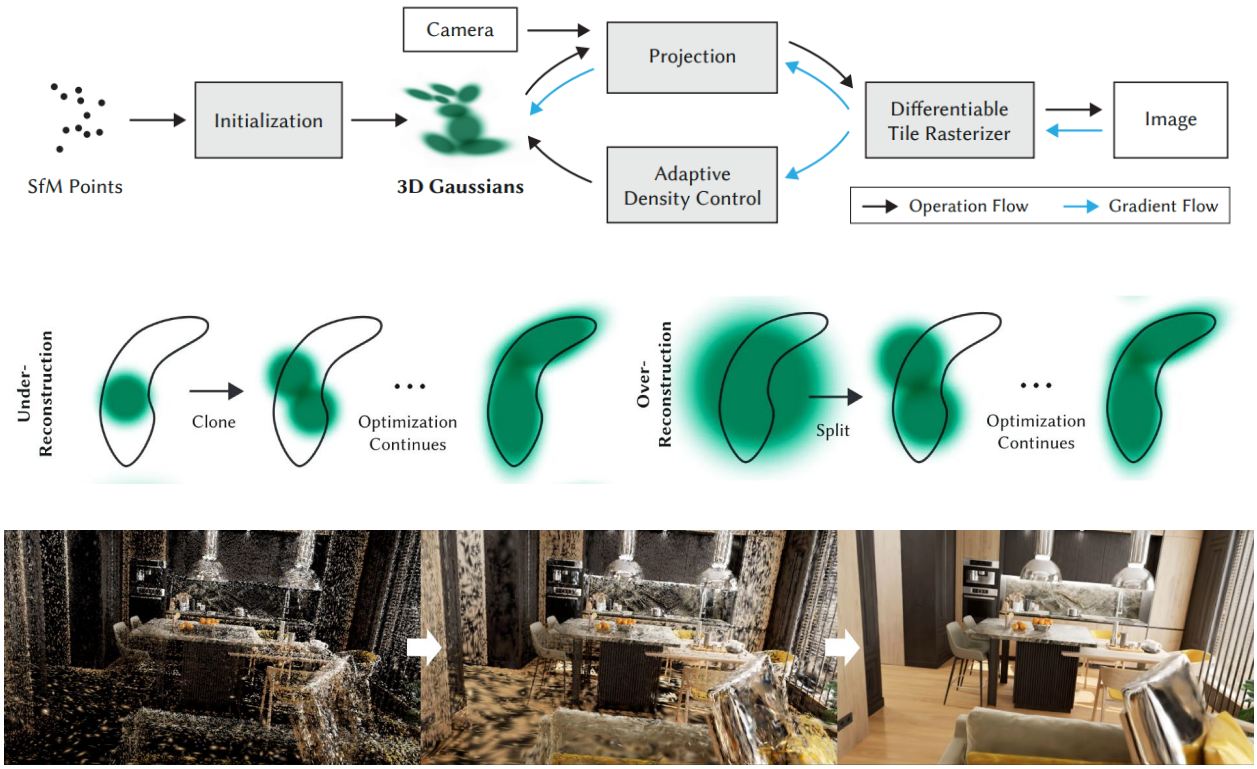


Fig 4. Ilustración *Gaussian Splatting*

Estado del arte:

- *Neural Radiance Fields (NeRFs)*

Es una técnica 3 años anterior a lo que consideraríamos los *Gaussian Splatting* actuales y por tanto tuvo una gran evolución durante ese tiempo acelerando considerablemente tanto los tiempos de entrenamiento como de inferencia especialmente con trabajos como *Instant NeRFs* de NVIDIA4 además de contar tanto con gran cantidad se soluciones tanto de código abierto como *NeRF Studio*, como comerciales y compatibles con los paquetes más usados de la industria como puede ser el ejemplo de *Volinga* con *Unreal*, *Pixotope* o *Disguise*.

Dado que tanto el entrenamiento como la inferencia de nuevos puntos de vista en *NeRFs* se realizan mediante redes neuronales profundas, el rendimiento puede estar limitado por la complejidad de la escena, el tamaño de renderizado necesario y los

recursos computacionales necesarios. Por otro lado, en algunos casos, *NeRFs* es capaz de alcanzar niveles de detalle extremadamente altos que no pueden ser alcanzados con otras técnicas.

- *Gaussian Splatting*

A pesar de ser una técnica más reciente (a partir de ahora siempre nos referiremos a su uso para simular *radiance fields*) la literatura al respecto estos días es inabarcable existiendo proyectos para mejorar su calidad, reducir su tamaño, habilitar su uso en grandes espacios, simulación de humanos, permitir animaciones o incluso para convertir estos gaussianos a polígonos. Aunque es ampliamente conocido, es conveniente recordar que estas investigaciones no siempre se terminan trasladando a un desarrollo listo para la producción. Además, por supuesto, existen tanto plataformas abiertas como comerciales que dan soporte a la generación de estos sistemas de una forma mucho más sencilla que la implementación de referencia.

Si bien en el momento de la creación del *Gaussian Splatter* hay bastantes similitudes con *NeRF* como la necesidad de un *"Structure From Motion"* para decidir el punto de vista original de cada imagen respecto a la escena inicial el uso del descenso del gradiente para optimizar los puntos de vista, en realidad no se trata de una red neuronal ni tanto en el entrenamiento ni mucho menos a la hora de renderizar la imagen. En cualquier caso, sería recomendable que quien tenga interés en conocer su funcionamiento interno consulte alguna de las guías de introducción a *Gaussian Splatting* como *A Comprehensive Overview of Gaussian Splatting* de Kate Yurkova.

Esto hace que tanto su velocidad de entrenamiento como su renderizado sea mucho más rápido e incluso con un mayor margen de mejora de cara al futuro.

Antes de hacer una comparativa entre ambos cabe destacar que, si bien ofrecen muchas ventajas frente a la representación poligonal clásica su capacidad, para ser editados (especialmente *NeRF*) o compatibles con elementos introducidos con por un artista como luces, reflejo, etc. está muy muy limitada y por tanto en muchos usos ninguno de los dos es una alternativa válida frente a las mallas 3D.

Similitudes:

- Representación de escenas 3D: los dos se utilizan para representar y renderizar escenas en espacios 3D permitiendo generar cualquier punto de vista.
- Utilización de datos 2D: Trabajan con datos de entrada en 2D (imágenes o puntos) para generar representaciones 3D.
- Ambos utilizan el descenso del gradiente para optimizar la representación en su entrenamiento.
- Usos: visualización médica, científica, VFX, animaciones, registro histórico etc.
- Están siendo desarrolladas técnicas *"in the wild"* tanto para *NERF* como para *Gaussian Splatting*, estas técnicas son más robustas ante variaciones del entorno como cambios de iluminación y pueden usar imágenes de entrenamiento no controladas abriendo así la puerta a combinar varias

fotografías públicas, turísticas o históricas para generar las escenas en lugar de requerir un *dataset* hecho a medida.

Diferencias:

- Los *NeRFs* utilizan redes neuronales tanto para aprender una función continua de la escena y como para renderizar la escena como un campo de radiancia, mientras que *Gaussian Splatting* usa nubes de puntos que contienen una funciones gaussiana y no depende de redes neuronales para la representación ni para el renderizado, siendo así compatible con sistemas tradicionales de rasterización y resultando en renderizados mucho más eficientes, llegando a ser posible ser usado en pequeñas escenas en dispositivos móviles a través de la web.
- Haciendo una simplificación, probablemente excesiva, podríamos decir que *NeRF* genera una red neuronal capaz de realizar imágenes 2d que corresponderían a cualquier punto de vista, mientras que *Gaussian Splatting* sí que genera unos datos 3D que han de ser interpretados realizando un trazado de rayos desde el propio punto de vista al objetivo hacia el que se mira.

Ventajas y desventajas:

NeRF

- Ventajas:
  - Bajo ciertas circunstancias puede tener mayor calidad en el detalle fino.
  - Tecnología más madura y menos sometida a cambios.
- Desventajas:
  - Proceso de entrenamiento costoso y lento.
  - Requiere recursos computacionales significativos.
  - Menos margen de mejora en el futuro y menos soportada en los paquetes de 3D al no tener una representación tridimensional.

Gaussian Splatting

- Ventajas:
  - Simplicidad y menor costo computacional.
  - Renderizado más rápido, adecuado para aplicaciones en tiempo real, con simplificaciones como la quantización puede llegar a ser ejecutado incluso en dispositivos móviles.
  - Menor tiempo de entrenamiento y mayores posibilidades de mejora.
  - Al tratarse de elementos en un espacio tridimensional puede combinarse con otra serie de elementos como mallas 3d.

  - A pesar de que su edición está muy limitada a borrar zonas aun así esto es una ventaja frente a la nula capacidad de editar lo que está dentro de una red *NeRF*.
- Desventajas:
  - Menor flexibilidad para representar escenas complejas con detalles finos.
  - Tecnología cambiante mes a mes.

Conclusión de la comparativa:

Si bien no se debe descartar *NeRF* de cara al futuro, e incluso en casos fuera de uso en otras industrias, como por ejemplo en la medicina, puede ser superior a *Gaussian Splatting* a día de hoy, parece claro que en la industria de la visualización este último es el claro ganador.

Esto además se ve reflejado en que muchos de los paquetes comerciales más conocidos como Unreal, Unity o After Effects ya están apostando por ello a través de *plugins* o paquetes *opensource* como Blender tienen sus propias integraciones hechas por terceros.

Así mismo, se están empezando a añadir pequeñas mejoras sobre el renderizado de *Gaussian Splatting* como cambiar la luz. Lógicamente esto se encuentra en un estado básico y no es útil para producción, especialmente por la falta de interacción a modo de sombras o reflejos entre las mallas 3d y los propios *splats,* pero indica que hay una gran cantidad de investigadores trabajando en mejorar estos sistemas y probablemente en un futuro tengan una integración muy mejorada.

Casos de Uso

Si bien no se trata de tecnologías que estén implementadas en el pipeline general de muchas empresas, está claro que siempre hablamos de tecnologías de reconstrucción de escenas 3D y por tanto vienen a sustituir a los anteriores métodos de creación de estas escenas.

Titularidad o Patentes

Puesto que no hablamos de tecnologías monolíticas, sino de una investigación inicial libre y reimplementaciones especializadas es necesario acudir a la fuente del origen en cada caso para consultar los términos y condiciones:

- Cine y Animación: generación de efectos visuales y escenas complejas en VFX y animación, permite la creación de gráficos por computadora realistas sin la necesidad de modelado manual detallado, ahorrando tiempo y recursos.
- Realidad Aumentada y Virtual (AR/VR): creación de entornos y objetos realistas para experiencias inmersivas mejorando la calidad visual en aplicaciones de AR/VR, ofreciendo una experiencia envolvente y realista.

- Simulaciones Médicas: creación de modelos 3D detallados de órganos y tejidos para simulaciones y entrenamientos médicos.
- Cualquier otra experiencia que requiera de objetos o entornos 3D como videojuegos, conservación del patrimonio, registro detallado de sets de rodaje, publicidad etc.

Disponibilidad y Costes de Implantación

De nuevo, no cabe una respuesta sencilla puesto que hay cientos de alternativas, desde las más sencillas de utilizar como Polycam, a implementaciones originales de la técnica sobre las que los investigadores pueden iterar para crear nuevas técnicas más óptimas. Asimismo, existen plataformas para su uso en local que no conllevan un coste económico en la licencia como *NeRF*studio, pero requieren un hardware costoso y otras que corren "en la nube" de forma que no necesitan este tipo de hardware pero sí una licencia mensual como Luma.

Además, existen *plugins* que permiten la inclusión de estos resultados en plataformas como Unreal, Unity, Blender o AfterEffects.

Limpieza de planos Asistido por IA Generativa

Introducción

La limpieza y preparación de planos de una producción de VFX incluye diferentes tipos de tareas en las que se incluye la eliminación de elementos del rodaje visibles, objetos anacrónicos y la corrección de reflejos y brillos entre otros.

Llevar a cabo estas tareas por plano, puede llegar a requerir una o dos jornadas de trabajo de un artista y es necesaria en la práctica totalidad de los planos de una producción de VFX, por esta razón, cualquier mejora en esta área tiene el potencial de generar un gran impacto positivo.

Los modelos de *inpainting* son herramientas de inteligencia artificial diseñadas para rellenar automáticamente partes faltantes o eliminar elementos no deseados de una imagen o video, completando la imagen de manera coherente. Estos modelos, basados principalmente en redes neuronales convolucionales y modelos de difusión, fueron introducidos por varias empresas de investigación en IA, como Stability, NVIDIA, Adobe y OpenAI, en los últimos 2 años. El uso de estos modelos de *inpainting* en combinación con los modelos de segmentación para generar máscaras puede ayudar a reducir el coste de las tareas de eliminación y limpieza de forma sustancial

Añadir a estos nuevos modelos como Lama, Stable Diffusion inpainting o herramientas como Photoshop Firefly o IOPaint a las ya disponibles como Nuke es la esencia de esta nueva técnica de limpieza de planos asistido por IA.

Descripción

La tarea de limpieza y preparación de planos de una producción de VFX incluye las siguientes actividades:

Nombre de la Tarea	Descripción
Eliminación de Aparejos	Eliminar equipos de filmación visibles como cables, aparejos o arneses
Eliminación de Objetos	Borrar digitalmente objetos no deseados del cuadro
Eliminación de Marcadores	Borrar marcadores de tracking usados como referencia para VFX
Limpieza de Sombras/Reflejos	Eliminar o ajustar sombras y reflejos de objetos removidos

Todas estas tareas tienen en común tres pasos:

1. Generar máscara: Seleccionar el área de la imagen que ocupan los elementos a eliminar o limpiar para generar una máscara
2. Generar nuevo contenido: Aplicar diferentes técnicas para generar un nuevo contenido que reemplazará el área seleccionada por la máscara eliminando o limpiando ciertos elementos
3. Composición del nuevo contenido en la imagen final: Aplicar técnicas de composición para integrar el nuevo contenido sin alterar el resto de la imagen de manera que se integre sin que sean visibles las costuras.

Estos tres pasos pueden ser asistidos por herramientas de IA y reducir así el tiempo necesario para ejecutarlas. Estas herramientas de IA se dividen en dos tipos:

- **Modelos pre-entrenados**: Estos modelos tienen la ventaja de no necesitar ser entrenados y solo requieren ser instanciados para llevar a cabo la tarea. Sin embargo, cuando la eliminación o sustitución de contenido de una imagen es muy específica puede que no generen la solución esperada.

  Ejemplos de este tipo de modelos son: Stable Diffusion en sus versión entrenada para *inpainting*, modelos como Brusnet o PowerPaint que convierten cualquier modelo de difusión en un modelo de *inpainting* o Lama que es un modelo entrenando para eliminar elementos de un imagen

- **Modelos entrenables**: Si somos capaces de generar una gran cantidad pares de ejemplos de la imagen con el contenido a reemplazar y la imagen con el

contenido reemplazado, se pueden utilizar modelos de redes neuronales como el usado por Nuke Copycat, para entrenar un modelo personalizado para un tipo específico de limpieza de imágenes. Estos modelos requieren la creación de los pares de imágenes de ejemplo para entrenar el modelo, pero a cambio se adaptan con precisión problema

A continuación, se describen las tres características más destacables de esta técnica, incluyendo qué modelos pre entrenados pueden ser utilizados ya que Nuke Copycat tiene una sección específica en el radar que detalla cómo usarlo para implementar esta técnica:

1. Aceleración de la generación de máscaras

Usando modelos como Segment Anything 1 o 2 se pueden generar máscaras de los objetos a eliminar o reemplazar con solo seleccionar uno o dos puntos de la imagen que forman parte de esa región evitando así que el artista tenga que definir el contorno de la región punto a punto o dibujándola con una brocha.

Fig 5. Ejemplo de una máscara (en amarillo) creada a partir de un punto (en verde) usando IOPaint

2. Aceleración de la eliminación de elementos de una imagen

Dentro de la familia de modelos entrenados para hacer *inpaint* existe un grupo especializado en específicamente en la eliminación áreas de una imagen como Lama, MAT o MI-GAN. Lama ofrece una buen equilibrio entre rapidez y calidad, MI-GAN es el modelo más rápido y que consume el menor número de recursos y MAT es el modelo más pesado y lento pero proporciona buenos resultados a nivel de calidad.

Fig 6. Resultado de aplicar Lama con la máscara de la imagen de la sección anterior. Una región de 256x256 pixels es eliminada y compuesta de nuevo en menos de 6 segundos usando una RTX A6000

3. Aceleración del reemplazo de una región de una imagen

Por último, los modelos más versátiles para reemplazar contenidos de una región de una imagen y componerlos de nuevo con la imagen completa con los modelos de difusión entrenados para hacer *inpainting*. Uno de los más utilizados es Stable Diffusion inpainting, pero recientemente se han lanzado modelos como Brushnet y PowerPaintv2 que pueden convertir cualquier modelo con la arquitectura de Stable Diffusion 1.5 en un modelo de *inpainting*.

Todos estos modelos tienen en común que esperan como *input* la imagen completa, una máscara que identifica el área a reemplazar y un texto que describe el nuevo contenido a generar. La versatilidad que ofrecen estos modelos a la hora de generar nuevo contenido e integrarlo en la imagen original es la que más potencial tiene de reducir costes y a mejorar la calidad de estos procesos.

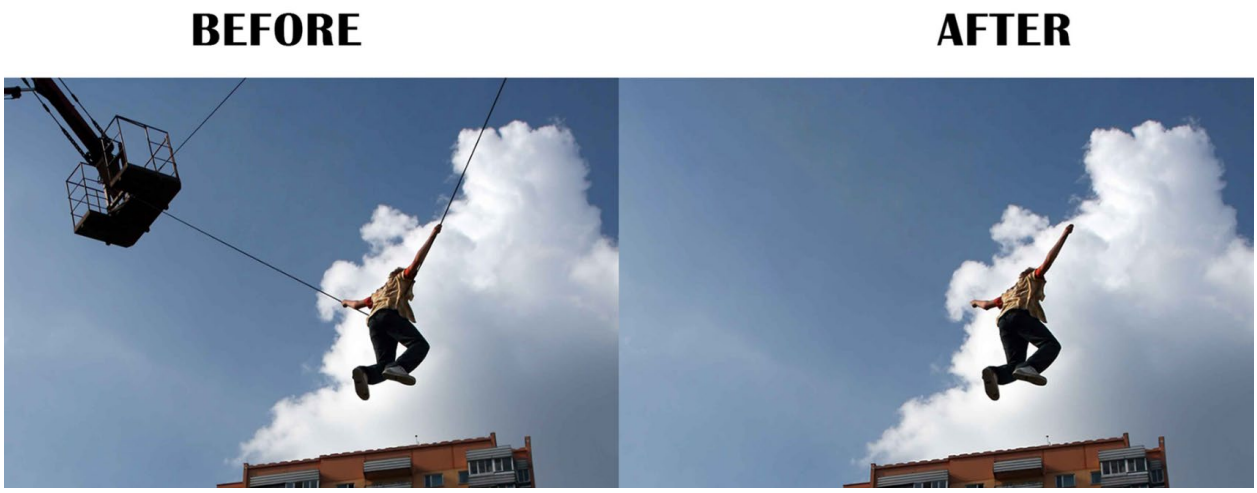


Fig 7. Ejemplo de un *frame* de un plano limpiado usando Segment Anything y Lama con IOPaint

Casos de Uso

Los modelos de *inpainting* son ideales para eliminar objetos pequeños y medianos en primer plano, corregir imperfecciones en superficies uniformes, y limpiar fondos desenfocados, donde pueden rellenar de manera natural utilizando información del entorno. Sin embargo, presentan limitaciones al enfrentarse a texturas complejas, detalles finos, zonas con reflejos y transparencias, interacción entre objetos en movimiento, características faciales humanas y escenas con iluminación dinámica, ya que en estos casos pueden generar resultados artificiales o incoherentes debido a la dificultad de replicar con precisión los detalles visuales y la dinámica de luz.

Casos de uso ideales

- Objetos pequeños y medianos en primer plano.

  Por ejemplo: Micrófonos, cables, pequeñas sombras o reflejos no deseados, o accesorios de producción que accidentalmente se han dejado en la escena.

  Estos objetos suelen ocupar un área relativamente pequeña de la imagen o video y no afectan en gran medida la estructura general o la coherencia de la escena. Los modelos de *inpainting* pueden llenar estas áreas vacías de manera eficiente utilizando la información circundante.

- Imperfecciones en superficies uniformes.

  Por ejemplo, grietas en paredes, manchas en el suelo, arrugas en la tela de fondo.Las superficies uniformes ofrecen patrones consistentes que los modelos de inpainting pueden aprender y replicar fácilmente. La eliminación de imperfecciones en estas áreas suele resultar en un acabado natural y convincente.

- Elementos en fondos desenfocados.

Por ejemplo: Personas o vehículos de fondo que están desenfocados, postes de luz, árboles o pequeños detalles en tomas de paisajes. Los fondos desenfocados tienen menos detalles, lo que facilita que el modelo rellene el espacio eliminado de manera coherente sin que se note la modificación.

- Logotipos o marcas no deseadas.
  Por ejemplo: Logotipos en ropa o en la utilería de la escena, señales de tráfico no deseadas. Eliminar estos elementos puede ser sencillo si no afectan de manera significativa la textura o patrón del entorno circundante.

Limitaciones

- Texturas complejas y detalles finos.
  Por ejemplo: Tramas de telas detalladas, complejas hojas de árboles en movimiento, patrones geométricos intrincados.
  Los modelos de *inpainting* pueden tener dificultades para replicar patrones complejos de manera precisa. Los detalles finos y las texturas complicadas requieren un análisis profundo y una capacidad de predicción detallada que los modelos actuales pueden no manejar perfectamente, resultando en artefactos o incoherencias.

- Zonas con reflejos y transparencias.
  Por ejemplo: Cristales, superficies de agua, objetos translúcidos.
  Los reflejos y las transparencias introducen variaciones complejas de luz y color. Replicar cómo se comportan estos reflejos de manera coherente puede ser complicado para los modelos de inpainting, lo que resulta en rellenos poco naturales o distorsionados.

- Interacción entre múltiples objetos en movimiento.
  Por ejemplo: Personas interactuando entre sí, escenas de acción donde hay múltiples elementos en movimiento rápido.
  En estas escenas, los cambios son dinámicos y pueden afectar la apariencia de los elementos eliminados. Mantener la coherencia espacial y temporal en este contexto es un reto significativo para los modelos de inpainting.

- Características faciales y detalles humanos.
  Por ejemplo: Expresiones faciales, detalles de ojos, boca, manos.
  Las facciones humanas son altamente detalladas y fácilmente reconocibles, lo que significa que cualquier imperfección en el inpainting es fácilmente identificable. Los modelos han mejorado mucho en este sentido, pero requieren técnicas de detailing o upscaling que son más costosas en tiempo de cálculo y en recursos de hardware.

Titularidad o Patentes

Las diferentes técnicas de uso de modelos de *inpainting* para la preparación y limpieza de planos no están patentadas y pueden usarse de forma comercial. Además, todos los modelos mencionados en esta sección tienen licencias de uso comerciales con algunas limitaciones que no impactan al uso de estas herramientas para la limpieza y preparación de planos de VFX.

Disponibilidad y Costes de Implantación

Los modelos de *inpainting* mencionados están disponibles bajo diferentes **licencias de código abierto** tipo MIT o Apace 2.0. Estas licencias permiten a las empresas implementar, modificar y distribuir la tecnología sin costo inicial, pero requieren recursos internos para su configuración y adaptación.

El único modelo que se ofrece comercialmente con un coste asociado es el de Copycat que viene integrado en Nuke, el software de composición desarrollado por The Foundry. Las diferentes licencias y costes se detallan en la sección específica de Copycat dentro del apartado de herramientas.

Infraestructura hardware necesaria

Para ejecutar modelos de *inpainting* de manera efectiva, especialmente en entornos empresariales que manejan grandes volúmenes de datos, es necesario contar con una infraestructura de hardware adecuada:

1. **GPU de alto rendimiento**: Los modelos de *inpainting* generalmente requieren GPUs (Unidades de Procesamiento Gráfico) potentes, como las de las series **NVIDIA RTX** que ofrecen la capacidad de procesamiento necesario para manejar operaciones de cálculo intensivo. Una configuración básica puede incluir una GPU con al menos 8-16 GB de VRAM, mientras que para operaciones más complejas o de gran escala se recomienda utilizar GPUs con 24 GB de VRAM o más.

2. **Procesador y memoria**: Además de la GPU, es esencial contar con un procesador de alto rendimiento (como un Intel i7 o superior) y al menos 32 GB de RAM para asegurar un rendimiento fluido y minimizar los cuellos de botella durante la ejecución de los modelos.

3. **Almacenamiento y red**: Un SSD de alta velocidad es fundamental para el manejo eficiente de datos y el acceso rápido a los modelos y archivos de entrenamiento. También se recomienda una conexión de red robusta para poder descargar los modelos que en ocasiones requieren hasta 20 GBs de parámetros.

Fuentes de ayuda y formación

Para asegurar una adopción exitosa de los modelos de *inpainting*, es crucial contar con acceso a fuentes de ayuda y formación. Al ofrecerse la mayoría de los modelos mencionados con licencia abierta la documentación y el soporte es proporcionada por la comunidad de usuarios a través de diferentes plataformas como HuggingFace, Github, Civitai, Discord y Youtube. Aquí se detallan las opciones disponibles de documentación y soporte

1. **Documentación oficial y tutoriales**: Herramientas como ComfyUI**,** Automatic111 **o** Fooocus, suelen ofrecer documentación detallada, tutoriales paso a paso y ejemplos de código que ayudan a los desarrolladores a comenzar rápidamente.

2. **Comunidad y soporte técnico**: Participar en foros de discusión en HuggingFace, Civitai, Discord, comunidades de GitHub, y grupos de usuarios específicos de IA y VFX puede proporcionar soporte práctico.

Herramientas

RunwayML

Introducción

Runway AI, Inc. (también conocida como Runway y RunwayML) es una empresa estadounidense con sede en la ciudad de Nueva York que se especializa en investigación y tecnologías de inteligencia artificial generativa. La compañía se centra principalmente en crear productos y modelos para generar videos, imágenes y diversos contenidos multimedia. Es especialmente reconocida por desarrollar los modelos comerciales de inteligencia artificial generativa de texto a video y video Gen-1, Gen-2 y Gen-3 Alpha.

Descripción

El producto Runway es accesible a través de una plataforma web y mediante una API como un servicio gestionado. Como parte de dicha plataforma incluye herramientas de edición de video como Remove Background que permite extraer objetos, personas o cualquier otro elemento de un clip de vídeo. Runway ofrece la capacidad de crear máscaras y extraer elementos de vídeos los siguientes pasos:

1. Importar clip de video. Es necesario subir los videos a editar a la plataforma de RunwayML en la nube.

2. Crear máscara. La herramienta permite generar máscaras del elemento a recortar mediante el posicionamiento de puntos de inclusión y de exclusión. Además, permite hacer un ajuste más fino de bordes complicados con una herramienta brocha de inclusión y exclusión. La máscara de recorte puede visualizarse en tiempo real en el frame elegido.
3. Aplicar la máscara a todo el vídeo. Tras terminar la generación de la máscara se puede usar la opción de vista previa para que la herramienta haga tracking de la máscara a lo largo de todo el video.

4. Exportar la máscara y el video. Runway proporciona múltiples opciones de formato y resolución.

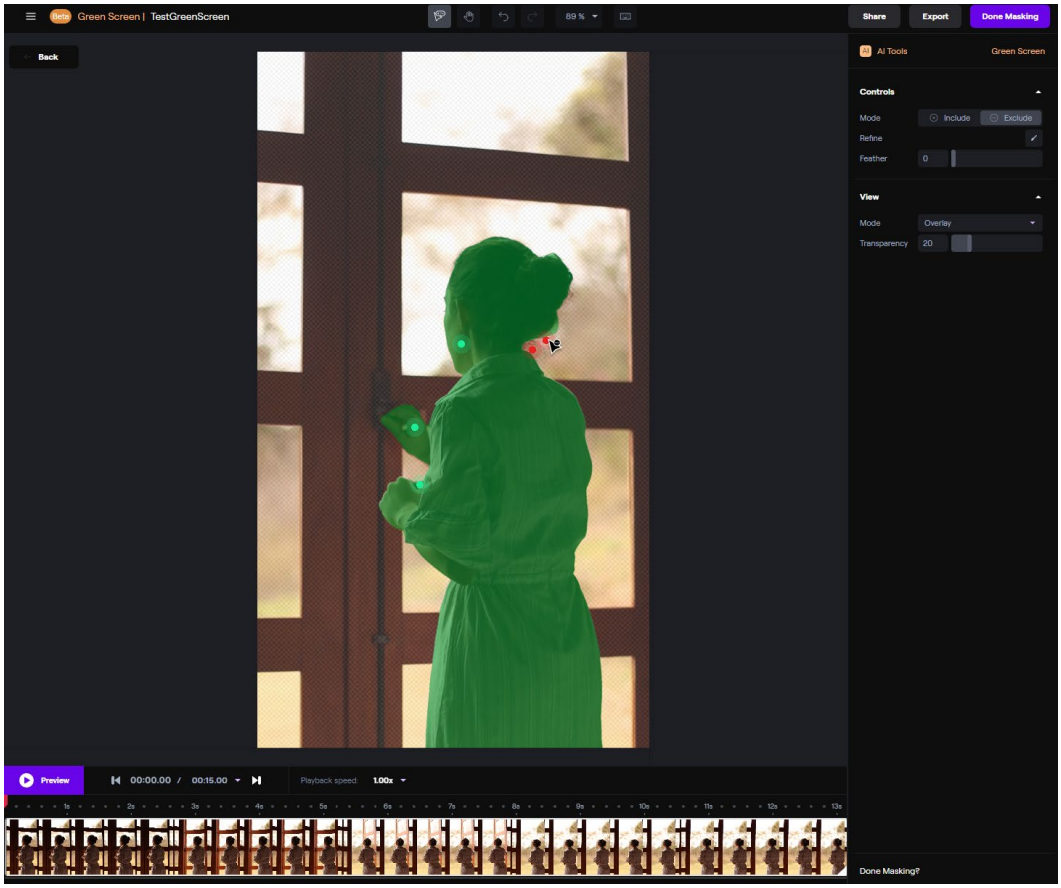


Fig 8. La imagen muestra una captura de la interfaz de usuario que proporciona Runway para generar máscaras de rotoscopia.

Casos de Uso

La herramienta ofrece sus mejores resultados con videos donde el elemento principal a recortar está bien definido con respecto al fondo, los movimientos de cámara son suaves. También es importante posicionar los puntos de inclusión en áreas que

identifiquen bien los diferentes colores y texturas que conforman el elemento a rotoscopiar.

Por el contrario, cuando el fondo consta de muchos elementos en movimiento, tiene poco contraste o el elemento a rotoscopiar sale y entra de la imagen, los resultados son decepcionantes y la calidad de la máscara no llega a la requerida para separar personajes principales del fondo. Tampoco proporciona herramientas eficaces para definir el contorno de superficies con pelo o poco definidas

Fig 9. La imagen muestra un ejemplo típico de oclusión del pantalón con una planta que define un borde poco definido y con mucho perímetro que resulta complicado identificar por la herramienta de RunwayML

Otro reto que plantea esta herramienta es la necesidad de transferir los videos a su plataforma en la nube antes de poder rotoscopiar. Esto puede hacer que ciertos estudios no puedan utilizarla ya que infringirían los acuerdos de seguridad de los datos y confidencialidad con sus clientes que les obligan a mantener todos los contenidos generados para un proyecto en soluciones de almacenamiento seguras e “in-situ”.

Titularidad o Patentes

Esta tecnología es propietaria de Runway AI, Inc. que la licencia por uso mensual o anual tal y como se detalla en las páginas web de términos y condiciones y su política de privacidad.

Disponibilidad y Costes de Implantación

Para hacer uso de esta tecnología solo es necesario contar con una buena conexión a internet y el pago de la licencia por uso mensual o anual tal y como se muestra en la página web donde comercializa este producto. La formación necesaria para el uso de esta herramienta es baja y Runway proporciona manuales y servicios de soporte adecuado al tipo de licencia y necesidades del cliente para permitir la adopción de la herramienta con facilidad.

SequoIA

Introducción

El desarrollo de herramientas ad-hoc específicamente diseñadas para tareas de rotoscopia en el sector audiovisual representa una solución avanzada y precisa para satisfacer las necesidades exigentes de la industria de la animación y los efectos visuales. Estas herramientas específicas, y por el momento escasas, deben estar diseñadas para ofrecer un rendimiento superior y una exactitud extrema en la creación de máscaras, superando las limitaciones de los modelos genéricos y comerciales.

Un ejemplo destacado de estas herramientas es SequoIA, desarrollada por Ilux Visual Technologies. Esta herramienta está diseñada para conseguir máscaras a nivel pixel, cumpliendo así con los más altos estándares de precisión requeridos por la industria. SequoIA ha sido rigurosamente testeada por empresas del sector de la animación y VFX, demostrando su capacidad para manejar tareas de rotoscopia complejas de manera eficiente y efectiva.

En la actualidad, las herramientas ad-hoc de segmentación diseñadas específicamente para el sector de la animación y los efectos visuales son escasas. Esta escasez se debe principalmente a su elevado nivel de especialización y a la complejidad inherente en su desarrollo e implementación. La creación de estas herramientas requiere una combinación de conocimientos técnicos avanzados en inteligencia artificial y un profundo entendimiento de las necesidades específicas del sector. Sin embargo, es razonable prever que, con el continuo avance de la tecnología y el creciente reconocimiento de los beneficios que estas herramientas pueden aportar, su disponibilidad y uso se incrementarán significativamente en el futuro próximo.

Descripción

Precisión: Uno de los mayores beneficios de este desarrollo ad-hoc de IA para tareas de rotoscopia es la capacidad de alcanzar un nivel excepcional de precisión y detalle. Esta herramienta está diseñada específicamente para generar máscaras a nivel de píxel, lo que permite capturar incluso los elementos más pequeños y complejos de

una escena. Esta exactitud es crucial en la producción de efectos visuales, donde cualquier error en la máscara puede afectar negativamente la calidad del producto final.

Sistema de Tracking Integrado: Una de las características más notables del sistema es la incorporación de un sistema de tracking integrado, que permite segmentar un objeto a lo largo de una escena sin necesidad de realizar este proceso fotograma a fotograma. Este sistema de seguimiento automático reduce significativamente el tiempo y el esfuerzo requeridos para mantener la coherencia de la máscara a través de múltiples fotogramas, mejorando así la eficiencia del flujo de trabajo.

Proceso Interno de Post-Procesado: El desarrollo incluye un sofisticado proceso interno de post-procesado que asegura la creación de máscaras de máxima precisión, a nivel pixel. Para cumplir este objetivo se utilizan una serie de técnicas avanzadas, las cuales son transparentes para el usuario y rápidas en su ejecución para la obtención de la máscara final. Este enfoque modular permite una mayor precisión y flexibilidad en la generación de máscaras, adaptándose a las variaciones y detalles minuciosos de los objetos en movimiento.

Supervisión del Operario: A pesar de las avanzadas capacidades automáticas de SequoIA, la supervisión del operario sigue siendo esencial para asegurar los mejores resultados posibles y minimizar errores. El operario tiene la posibilidad de realizar ajustes y correcciones en tiempo real, garantizando que la máscara final cumpla con los estándares de calidad requeridos. Esta combinación de automatización y supervisión humana optimiza tanto la precisión como la eficiencia del proceso de rotoscopia.

Personalización: Este desarrollo ad-hoc ofrece también un alto grado de personalización, permitiendo ajustar los algoritmos y procesos según las necesidades específicas de cada proyecto. Esta flexibilidad facilita la adaptación a diversos tipos de contenido y requisitos de producción, proporcionando una solución a medida que puede ser optimizada para diferentes escenarios y tipos de archivos.

Casos de Uso

Casos de uso adecuados

- Objetos de alto contraste. En aquellas escenas donde los objetos a segmentar tienen un contraste notable con respecto al fondo SequoIA puede identificar y seguir con precisión los bordes de los objetos, facilitando la creación de máscaras exactas
- Escenas con movimientos no bruscos. Los sistemas de IA que conforman SequoIA pueden seguir eficazmente los movimientos fluidos, como el caminar de un personaje o el movimiento de objetos en una trayectoria estable. Este tipo de movimiento permite a SequoIA aplicar máscaras precisas y consistentes, reduciendo el esfuerzo manual.
- Recursos visuales de alta resolución. SequoIA está optimizada para manejar recursos visuales de alta resolución, permitiendo trabajar con detalles finos y

ofreciendo resultados de alta calidad. Esto es crucial en producciones de cine y televisión donde la nitidez y precisión son fundamentales.

- Creación de múltiples máscaras en una misma escena. Este sistema puede gestionar y crear múltiples máscaras dentro de una misma escena, permitiendo una edición eficiente y coherente. Esto es útil en escenas complejas donde varios elementos necesitan ser aislados y manipulados individualmente.

Limitaciones

- Escenas con movimientos abruptos. Los sistemas de IA para tracking y segmentación pueden experimentar dificultades para seguir correctamente los contornos en movimientos rápidos, en estos casos el sistema necesitará de una mayor supervisión por parte del usuario.
- Ambientes con oclusiones frecuentes. Al trabajar con escenas en presencia de oclusiones, el sistema puede no ser capaz de identificarlo correctamente al reaparecer, necesitando de una mayor guía del operario.
- Escenas con bajo contraste. La tecnología de IA en SequoIA requiere puntos característicos bien definidos para funcionar correctamente. En superficies uniformes o con bajo contraste, puede fallar en detectar y seguir el movimiento con precisión. No obstante, al tratarse de un desarrollo ad-hoc, y permitir trabajar con archivos EXR, el usuario puede cambiar el contraste para facilitar el trabajo del sistema

Titularidad o Patentes

SequoIA es una tecnología propietaria desarrollada internamente por la empresa Ilux Visual Technologies. No se encuentra disponible públicamente en la web; se ofrece exclusivamente como servicio ad-hoc para cada cliente.

Licencia:

El uso de SequoIA está regido por un acuerdo de licencia específico para cada cliente que se proporciona al iniciar el servicio. Este acuerdo define los términos de uso, las responsabilidades de ambas partes y las restricciones aplicables.

Políticas de Privacidad:

- **No recopilación de datos**. Al ser SequoIA una solución que se ejecuta de forma local en los equipos del cliente, no se recopila, almacena ni procesa ningún tipo de dato generado o utilizado por el sistema.
- **Datos Locales**. Todos los datos, incluidos los archivos de vídeo y las configuraciones del proyecto, se manejan exclusivamente en el entorno local del cliente.

- **Control completo de los datos.** Los clientes tienen control total sobre los datos procesados por SequoIA en sus sistemas locales. Esto incluye la capacidad de modificar, eliminar y transferir dichos datos según sea necesario.
- **Soporte técnico.** La empresa ofrece soporte técnico para ayudar con la instalación, configuración y uso de SequoIA. Este soporte se proporciona respetando la privacidad y seguridad de los datos del cliente.

Propiedad Intelectual:

Este software está protegido por derechos de propiedad intelectual. La protección de estos derechos asegura que SequoIA se mantenga como una solución exclusiva y avanzada, permitiendo a los usuarios beneficiarse de su tecnología innovadora de manera segura y legítima.

Disponibilidad y Costes de Implantación

SequoIA es una herramienta avanzada para rotoscopia en VFX y animación, diseñada para ofrecer un soporte integral y personalizado a cada cliente. Debido a la complejidad y especificidad de las necesidades de cada proyecto, SequoIA no dispone de licencias predefinidas ni precios fijos. En su lugar, se trabaja de manera personalizada con cada cliente para determinar la mejor configuración y costo basado en sus requerimientos específicos. En esta evaluación personalizada se incluye un análisis de las necesidades del cliente, un estudio de la solución a medida y una determinación final de costes.

Dada su naturaleza compleja y las variadas necesidades de los clientes, no existe un estándar único de hardware recomendado. La infraestructura necesaria se determina caso por caso, teniendo en cuenta las necesidades y factores específicos del proyecto. Para ello, se realiza una evaluación personalizada de infraestructura, donde entran en consideración factores como:

- Análisis de la arquitectura computacional actual. Se evalúa la infraestructura de hardware existente para asegurar la compatibilidad y determinar si se requieren actualizaciones o mejoras
- Determinación del flujo de procesado. Se analiza también el flujo de trabajo del cliente, incluyendo la cantidad de datos procesados, la frecuencia de uso y los requisitos de rendimiento. De esta forma se consigue optimizar SequoIA para el flujo de trabajo específico de cada cliente, asegurando una operación fluida y eficiente
- Evaluación del tipo de contenido. Se considera el tipo de contenido que se procesará, como la resolución del video, la complejidad de las escenas y los efectos visuales requeridos. De esta forma se pueden proveer recomendaciones de hardware que permitan manejar eficientemente el contenido específico del cliente.

LUMA AI Interactive Scenes

Introducción

Luma AI es una compañía estadounidense fundada en 2021 en San Francisco, se desarrolla software de inteligencia artificial para la industria de los videojuegos, inmobiliaria y de comercio electrónico.

En su portfolio de soluciones está Capture, un software de captura de entornos 3D a partir de imágenes 2D. Para esta tarea usa algoritmos NeRF y *Gaussian Splatting*.

Descripción

Luma AI Capture es accesible a través de una plataforma web, una aplicación móvil de iOS y Android. Como parte de dicha plataforma incluye también herramientas para visualizar los entornos 3D capturados y para generar mallas 3D con texturas de dichos entornos.

Luma AI Capture recibe como entrada una serie de imágenes 2D que muestran un objeto o entorno desde múltiples puntos de vistas y que tengan amplias áreas de solapamiento entre imagen e imagen. Cuantas más imágenes se suministren y desde más puntos de vista mejor será la recreación generada por Luma AI Capture.

Para generar dichas imágenes se puede usar la aplicación móvil de Luma AI o una cámara de fotos convencional o una cámara 360. Posteriormente las imágenes tienen que transferirse a la plataforma en la nube de Luma AI donde son procesadas de forma asíncrona hasta que el modelo *NeRF* o la representación con *Gaussian Splatting* ha sido calculado. En ese momento el usuario es notificado y a través de una aplicación web puede tener acceso a la captura para visualizar o para transferirse a su ordenador el modelo PLY de *Gaussian Splatting* o la geometría 3D de la captura en formato GLTF, USDZ o OBJ.

Luma AI también proporciona un *plugin* para Unreal 5 para visualizar en entorno de forma interactiva.

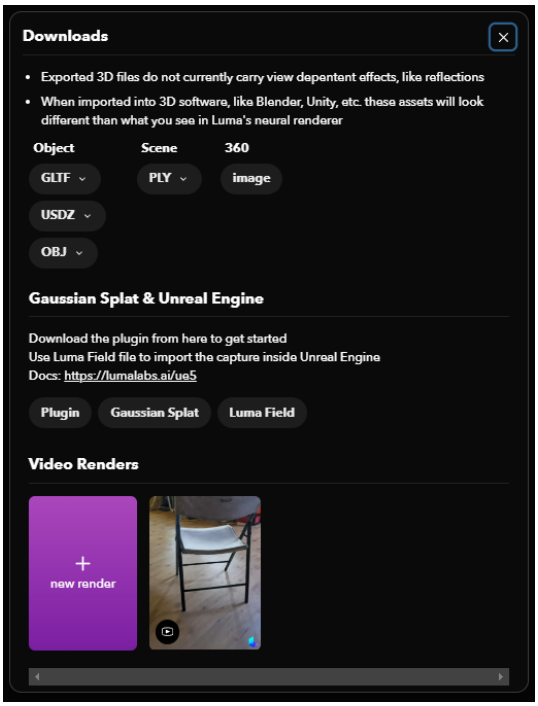


Fig 10. Diferentes opciones disponibles para exportar el entorno capturado

Casos de Uso

Luma AI Capture es particularmente sencillo de utilizar desde las aplicaciones móviles y sin ser un experto en fotogrametría permite capturar objetos y entornos de manera fácil y efectiva. Además, la calidad aumenta se usa un equipamiento de fotografía más profesional y una metodología de captura más estructurada. Esto dota a la solución de una versatilidad muy interesante para los supervisores de VFX durante los rodajes permitiéndoles capturar escenarios grandes con equipamiento profesional y pequeños objetos o salas con dispositivos móviles.

Sin embargo, la plataforma de Luma AI capture no proporciona herramientas para editar la captura y la geometría 3D generada solo es útil como referencia y no puede utilizarse directamente una escena 3D de producción ya que la mallas de triángulos no están cosidas y las texturas son muy heterogéneas.

Otros datos de interés sobre LUMA AI

- Incluye una guía de buenas prácticas sobre captura, que es útil no solo para esta plataforma si no es aplicable a cualquiera. https://docs.lumalabs.ai/MCrGAEukR4orR9
- Asimismo, también ofrece de forma gratuita y de código abierto una librería para ThreeJS que permite visualizar *Gaussian Splatting* en la web: https://lumalabs.ai/luma-web-library
- La empresa parece estar prestando cada vez menos atención a este producto y se está focalizando más a Dream Machine, un generador de video a partir de texto (y en el futuro a partir de imágenes y/o video) https://lumalabs.ai/dream-

machine y el enlace a la parte de captura cada vez está más oculto, así que el futuro de LUMA AI Capture no está muy claro.

Titularidad o Patentes

Luma AI Capture es una tecnología propietaria de Luma AI y la ofrece a terceros comercialmente con los siguientes términos y condiciones y con las siguientes condiciones de privacidad

Disponibilidad y Costes de Implantación

Actualmente Luma AI Capture está disponible de forma gratuita con un límite de 5Gb por proyecto.

Admite una gran variedad de formatos de salida

- Geometría GLTF, USDZ, OBJ, PLY
- Imagen: proyección 360º en formato .jpg
- *Gaussian Splatting*: formato .ply estandarizado para *Gaussian Splatting* y formato propio .luma compatible con su plugin para Unreal Engine 5

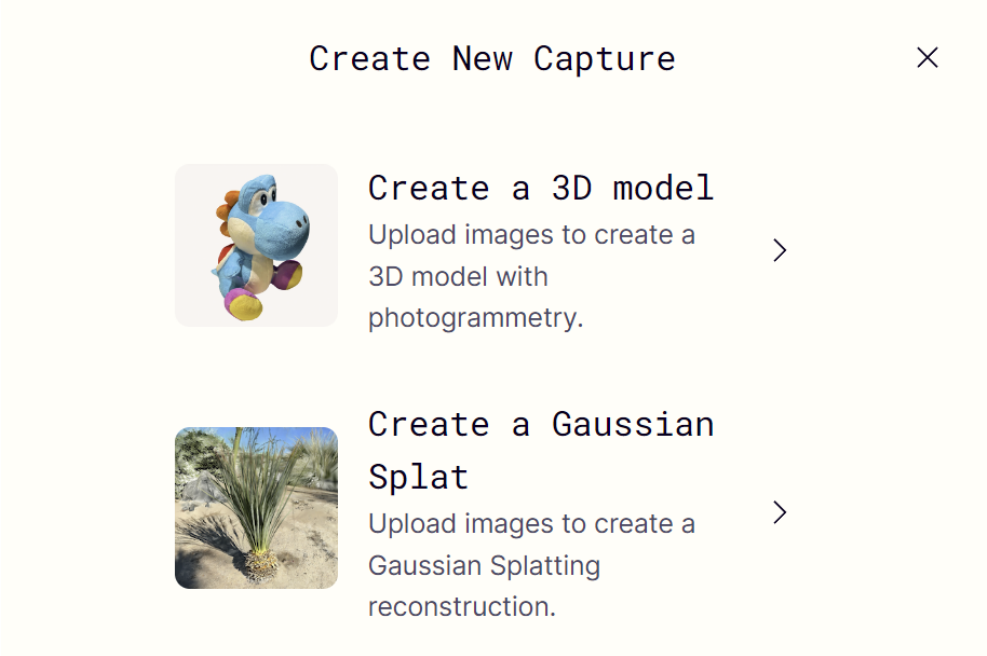
Polycam

Introducción

Si bien Polycam comenzó en el año 2020 como una aplicación para iPhone y iPad que te permitía escanear objetos y espacios en 3D utilizando la tecnología LiDAR del dispositivo a día de hoy en su web vemos distintos módulos entre ellos Photogrammetry que permite usar un set de fotos tomadas anteriormente o un video para generar mallas tridimensionales o *Gaussian Splatting*. En este documento sólo nos referiremos a las características del módulo *Photogrammetry*.

El equipo detrás de Polycam se encuentra en el área de San Francisco y provenían de otra empresa llamada Ubiquity6 que fue adquirida por Discord.

Descripción



El uso a través de web es trivial y simplemente requiere subir un set de imágenes que la plataforma convertirá a una malla 3d o un *Gaussian Splatting*. En su versión gratuita está limitado a 50 imágenes y si bien permite su visualización online no permite la descarga de los resultados.

Casos de Uso

Fig 11. Ejemplo de generación con Polycam

Se trata de una herramienta que permite generar fotogrametría y mallas 3d, o *Gaussian Splatting* online de una forma totalmente trivial permitiendo en su método de pago descargarlo en distintos formatos para ser usado posteriormente en otras aplicaciones.

Si bien a menudo su calidad puede ser inferior a Luma en determinadas ocasiones donde esta no es capaz de alinear los puntos de vista correctamente Polycam resuelve la escena correctamente dando un resultado claramente superior.

Titularidad o Patentes

La licencia comercial completa puede consultarse en https://poly.cam/terms_and_conditions.pdf

Disponibilidad y Costes de Implantación

Se trata de un servicio en la nube con tres niveles de acceso dependiendo de la cantidad pagada: https://poly.cam/pricing pero su uso es increíblemente sencillo.

Volinga

Introducción

De nuevo nos encontramos con otra herramienta orientada a la creación de NERF y *Gaussian Splatting*. Esta empresa española fundada en 2022 y abrió su tecnología al público en junio de 2023 con un enfoque inicial exclusivamente centrado en NERF.

Volinga ha construido gran parte de su sistema sobre Splatfacto, un proyecto de código abierto que utiliza *NeRF*studio en su núcleo. La empresa no solo utiliza esta

tecnología, sino que también contribuye activamente al mantenimiento y desarrollo del código de Splatfacto.

A diferencia de otras plataformas similares más generales, Volinga se distingue por su enfoque especializado en proyectos de producción virtual, ofreciendo soluciones optimizadas para este tipo de entornos

Descripción

Aunque algunas partes del pipeline, como el registro de cámaras, aún no alcanzan el nivel de precisión de herramientas como Polycam o Luma AI, Volinga se centra en los sistemas de producción virtual haciendo que que sea la plataforma que más y mejores características tenga en sus plugins para sistemas de este sector como son Unreal Engine, Disguise y Pixotope.

A diferencia de las soluciones aportadas por otras plataformas los plugin de Volinga permiten hacer cierto nivel de edición sobre los archivos una vez importados con características como:

- *Crop Volume Component*: permite seleccionar zonas que pueden ser incluidas o excluidas del renderizado.
- *Post-processing settings* que permite variar el color del escenario capturado dentro de los propios plugins sin volver a regenerar la captura.
    - *Tint*: tintado con un color sólido.
    - *Saturation*: cambio de saturación.
    - *Shadows*: Cambio de color en las bajas sombras.
    - *Gamma*: Cambio de color en los tonos intermedios.
    - *Lights*: Cambio de color en las altas luces.
    - *Hue*: Cambio de tono en la rueda de color hue.

- Características experimentales, se trata de nuevas posibilidades que si bien aún no están garantizadas en su uso en producción añaden un nuevo nivel de control sobre las capturas:
    - *Additive lighting*: Permite una interacción parcial de la captura con las luces direccionales.
    - *Normal Generation Method*: Volinga incluye 3 métodos para calcular las normales de la superficie que se usará para calcular el como la luz afecta al objeto.
    - *Ambient Light*: las sombras duras generadas por las luces direccionales.

Casos de Uso

Dado su especialización en sistemas de producción virtual su uso más obvio es ser usado para producciones de este tipo bien sean de cine, series o programas de TV,

pero por otro lado al disponer de un plugin muy optimizado para Unreal Engine también podría ser una excelente opción para sistemas de AR de alto nivel o videojuegos.

Titularidad o Patentes

Cuando hablamos de Volinga, sin duda, su fuerte reside en los plugins que dan soporte a las plataformas de producción virtual.

Junto con dichos plugin también sería destacable su formato .nvol que sustituiría a otros más comunes en el sector de *Gaussian Splatting* como .ply y .gsplat. Este formato es un contenedor tanto para *NeRFs* como para *Gaussian Splatting* siendo un contenedor muy optimizado para ambos formatos.

Disponibilidad y Costes de Implantación

Costes de la licencia https://volinga.ai/pricing según tipo de usuario:

- *Freemium*: Acceso limitado, permite hasta 5 conversiones mensuales y 3 archivos ".nvol" generados. No se permite el uso comercial y el soporte se brinda a través de foros comunitarios, además el contenido generado contiene marca de agua.
- *Pay on Demand*: 50€ de coste por creación con descuentos por volumen de compra (185€ por 5 o 250€ por 10 capturas).

- *Enterprise*: Plan personalizado para empresas, con soporte completo y opciones de uso comercial bajo acuerdo específico, número de capturas ilimitado que se generan en la máquina local del cliente en lugar de en los servidores de Volinga.

Dado que se trata de un producto para su uso profesional especialmente en plataformas como Unreal Engine se puede dar por hecho que el hardware capaz de correr dichas plataformas será suficiente para ejecutar Volinga.

Si hablamos de las fuentes a las que el usuario puede acudir podríamos destacar tanto la página de documentación https://volinga.ai/docs/index.html como su canal de discord para ayuda entre usuarios de la comunidad https://discord.com/invite/XxVdfWvcge Por otro lado los clientes del plan Enterprise pueden solicitar ayuda profesional por correo electrónico al departamento de atención al cliente.

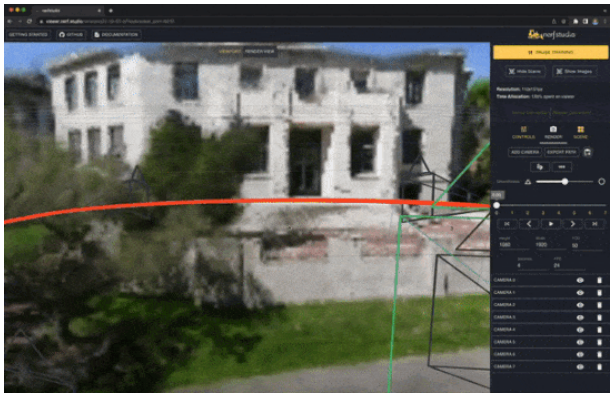
El uso tanto de la plataforma como de los plugin es intuitivo y sencillo, pero como es lógico para ser usado en producción virtual es necesario contar con personal artístico capaz de llevar a cabo ese tipo de producciones.

NeRFstudio

Introducción

*NeRFs*tudio es una herramienta de código abierto y multiplataforma creada en 2022 por un peque grupo de investigadores del campo de la inteligencia artificial y gráficos por ordenador, pero actualmente está mantenida por 233 personas y empresas que contribuyen sin ánimo de lucro. Originalmente fue diseñada para trabajar con Neural Radiance Fields (*NeRFs*), pero actualmente ya soporta *Gaussian Splatting* e incluso SDFs (Signed Distance Fields).

Descripción



Soporte para *NeRFs* y *Gaussian Splatting*:

- *NeRFs*: Modela la radiancia en un espacio 3D utilizando redes neuronales profundas, permitiendo la síntesis de vistas realistas desde diferentes ángulos.
- *Gaussian Splatting*: Técnica que usando distribuciones gaussianas en el espacio permite representaciones más eficientes y rápidas, mejorando a menudo la calidad y la velocidad frente a NeRF.

Coste y privacidad:

- Al ejecutarse en el propio ordenador del usuario no incurre en costes de terceros ni a nivel de licencia ni de servicios en la nube, asimismo los datos no salen del control del usuario y por tanto se garantiza la privacidad sobre los mismos.

Interfaz:

- Frente a la mayor parte de las soluciones opensource *NeRFs*tudio tiene una interfaz que permite al usuario interactuar sin acudir a la linea de comando ni necesidad de conocimientos de programación (más allá de la propia instalación que necesita conocimientos básicos de Python)

Comunidad y soporte:

- Cuenta con el apoyo de una comunidad activa de investigadores y desarrolladores, que ofrece documentación, tutoriales y canales de contacto para compartir avances y resolver duda como su canal oficial de Discord que actualmente tiene más de 5000 miembros activos https://discord.com/invite/uMbNqcraFc

Casos de Uso

Fig 12. Ejemplo de uso de NeRFStudio

Si bien *NeRFs*tudio no es la herramienta más optima en cuanto a calidad se refiere donde otras alternativas comerciales como Polycam o Luma son mejores su licencia permisiva y su uso en la propia máquina del usuario lo hacen una herramienta perfecta para casos como:

- Prototipado: ideal para los momentos en los que el usuario solo quiere comenzar con este tipo de técnicas sin necesidad de amplios conocimientos técnicos y sin incurrir en costes.
- Investigación: dado que permite iterar entre varios modelos de reconstrucción es una herramienta ideal para comparar técnicas o desarrollar nuevos desarrollos que impulsen la creación y uso de NeRF y/o *Gaussian Splatting*.
- Visualización: permite abrir entrenamientos generados con otras plataformas y visualizarlos en la máquina del usuario.
- Privacidad: frente a las plataformas en la nube, la seguridad de los datos está garantizada y por tanto puede ser la solución frente a proyectos que exigen confidencialidad estricta.

En conclusión, *NeRFs*tudio es ideal para usuarios que buscan una solución equilibrada entre facilidad de uso, funcionalidad avanzada y rendimiento, especialmente en entornos de investigación, educación y desarrollo de prototipos rápidos. Sin embargo, para proyectos que requieren personalización extrema, integración profunda con sistemas existentes, o producción a gran escala, otras soluciones pueden ser más adecuadas.

Titularidad o Patentes

*NeRFs*tudio está licenciado bajo Licencia Apache que permite su uso libre incluso en obras derivadas y proyectos comerciales: https://github.com/*NeRFs*tudio-project/*NeRFs*tudio?tab=Apache-2.0-1-ov-file#readme

Disponibilidad y Costes de Implantación

*NeRFs*tudio es una herramienta de código abierto disponible en plataformas como GitHub bajo licencia Apache, lo que significa que no requiere ningún tipo de pago, incluso para proyectos comerciales. Sin embargo, aunque su uso es intuitivo, su instalación requiere conocimientos básicos de Python, por lo que un artista puede necesitar asistencia para ponerla en marcha.

A diferencia de las plataformas en la nube que realizan la inferencia en servicios de terceros, todos los cálculos en *NeRFs*tudio se ejecutan en la máquina del usuario. Esto implica que se necesita un hardware más potente, incluyendo una tarjeta gráfica de alta gama compatible con CUDA.

Para la formación sobre el uso de la herramienta, además de la documentación oficial disponible en la página de ayuda de *NeRFs*tudio, existen numerosos tutoriales tanto escritos como en video disponibles en la red. También está el canal de Discord de la comunidad de *NeRFs*tudio, accesible en este enlace, que ofrece soporte y recursos adicionales.

Nuke Copycat

Introducción

CopyCat es una herramienta avanzada disponible en las versiones NukeX y Nuke Studio, ambas pertenecientes al software de composición digital Nuke, desarrollado por Foundry, una empresa británica reconocida en la industria de los efectos visuales (VFX) y la postproducción. Lanzada en 2021, CopyCat permite copiar efectos específicos de una secuencia, como *garbage matting*, reparaciones de belleza o reducción de desenfoque, a partir de un pequeño número de fotogramas seleccionados.

A través del uso de aprendizaje automático, y con tan solo unos pocos datos de entrenamiento, CopyCat permite a los usuarios entrenar una red personalizada para replicar estos efectos en toda la secuencia de manera eficiente. Esta herramienta es utilizada por artistas de efectos visuales y profesionales de postproducción para automatizar y optimizar tareas repetitivas, manteniendo altos estándares de calidad visual en sus proyectos.

Descripción

Copycat es una funcionalidad de Nuke que permite a los artistas crear modelos de IA personalizados directamente dentro de Nuke, potenciando la capacidad de automatizar tareas repetitivas y mejorar la eficiencia en la producción de VFX. La red neuronal utilizada en Copycat usa una arquitectura de Multi-Scale Recurrent Networks diseñada para manejar datos espacio-temporales con diferentes escalas de resolución. Esta red necesita ser entrenada de forma supervisada lo que implica generar un conjunto de imágenes de entrenamiento y posteriormente ejecutar las múltiples iteraciones de aprendizaje hasta que la red consigue editar las imágenes tal y como esperamos. Por ejemplo, si queremos entrenar Copycat para que identifique matrículas de coches y genere una máscara sobre las matrículas, entonces tendríamos que generar múltiples imágenes donde se puedan ver matrículas de coches y sus correspondientes máscaras. Una vez entrenado el modelo, se pueden guardar sus parámetros y usarse para generar imágenes en grafos de Nuke como cualquier otro nodo.

A continuación, se describen sus características más relevantes:

- **Integración directa en el Flujo de Trabajo de VFX**

  Nuke Copycat se integra sin problemas en el pipeline de VFX, permitiendo a los artistas aplicar sus modelos de IA entrenados en tareas como rotoscopia, eliminación de objetos, y mejoras en la calidad de imagen. También permite usar la funcionalidad de Nuke para generar de forma procedural las imágenes que se usarán para entrenar el modelo. Además se pueden usar los nodos de Nuke para implementar estrategias de *data augmentation* para que a partir de un pequeño número de imágenes de ejemplo, se puedan generar cientos de imágenes de entrenamiento.

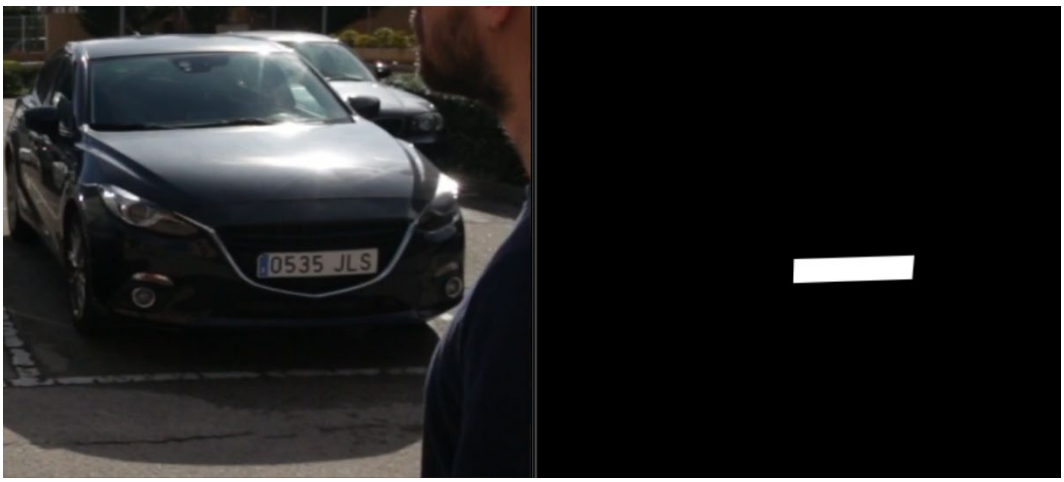


Fig 13. Ejemplo de una pareja de imágenes de entrenamiento para que Copycat aprenda a rotoscopiar matrículas de coches.

- **Entrenamiento de Modelos de IA Personalizados**

Permite a los usuarios entrenar modelos de IA específicos para sus necesidades, directamente dentro de Nuke, sin necesidad de salir del entorno de composición. En muchas ocasiones, ciertas producciones de VFX plantean retos únicos que son difíciles de resolver con modelos pre entrenados. Usando Copycat se pueden generar modelos que se adaptan mejor y en menos tiempo a problemas específicos de la industria.

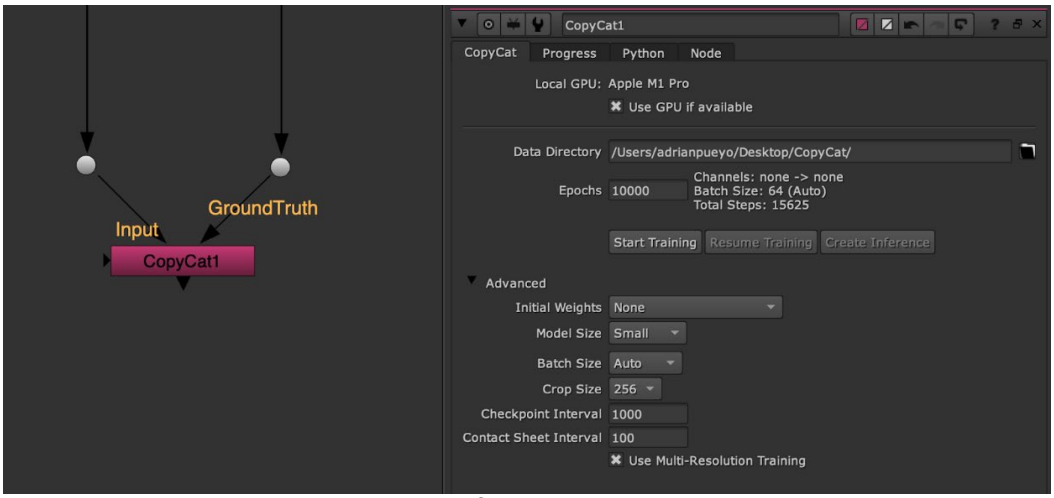


Fig 14. Vista de los parámetros de configuración del proceso de entrenamiento de la red neuronal.

- **Automatización y Escalabilidad**

Facilita la automatización de procesos repetitivos, reduciendo significativamente el tiempo de producción y permitiendo a los equipos de VFX escalar sus operaciones con mayor eficiencia. Una vez entrenado el modelo y verificada su correcto funcionamiento, se procede a guardar sus parámetros y a partir de ese momento se puede utilizar el modelo en cualquier grafo de Nuke como un nodo más permitiendo así aplicar la solución generada por la red neuronal a múltiples planos en una producción.

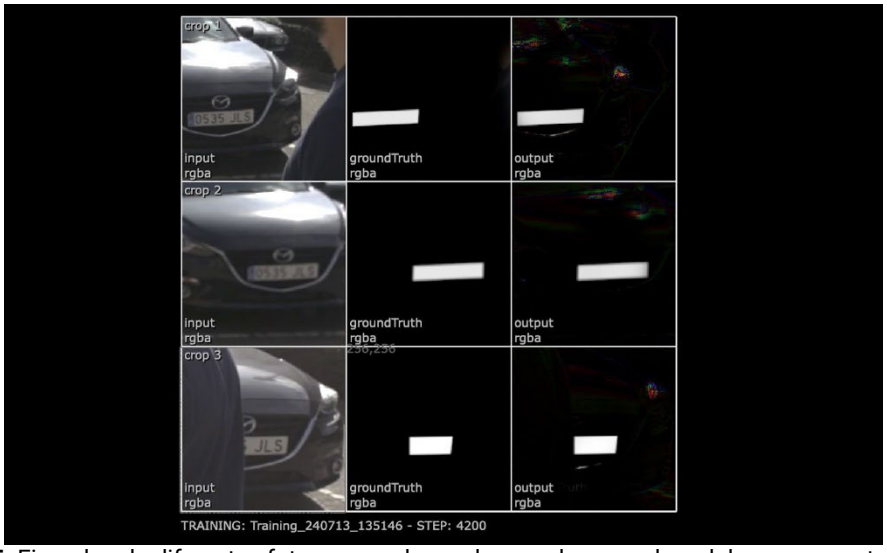


Fig 15. Ejemplos de diferentes fotogramas de un plano en los que el modelo una vez entrenado es capaz de generar una máscara muy similar a la generada manualmente por un artista.

Casos de Uso

A continuación, se detallan algunos de los casos de usos principales de Copycat, acompañados de ejemplos para ilustrar las habilidades de la herramienta.

- **Reparaciones de Belleza.**

  Nuke Copycat se ha convertido en una herramienta indispensable para reparaciones de belleza, permitiendo realizar retoques precisos y automáticos en las imágenes, optimizando el tiempo de postproducción sin sacrificar calidad.

Fig 16. Fotogramas de una de las secuencias de Dance First, proporcionadas por UFX Studios, donde se puede ver como gracias a Copycat se llevó a cabo la limpieza de la herida que tiene el actor en el ojo.

- **Creación de Máscaras y Modificaciones**

  Con Nuke Copycat, es posible entrenar el modelo para generar máscaras de colores específicas asociadas a objetos concretos dentro de una escena. Este proceso implica alimentar al modelo con ejemplos donde se etiqueten manualmente los colores y objetos deseados. A medida que el modelo aprende a identificar estas características, puede automatizar la creación de máscaras precisas que delimitan cada objeto según su color. Estas máscaras pueden luego utilizarse para aplicar modificaciones posteriores, como cambios de color, ajustes de iluminación o efectos visuales específicos, mejorando así la eficiencia y precisión en la postproducción. A continuación, se muestran diversos ejemplos de este proceso de creación de máscaras, y posteriores modificaciones, proporcionados por Edu León, Senior Digital Compositor para VFX.

Fig 17. Creación de máscaras de colores por medio de CopyCat en diferentes clips.

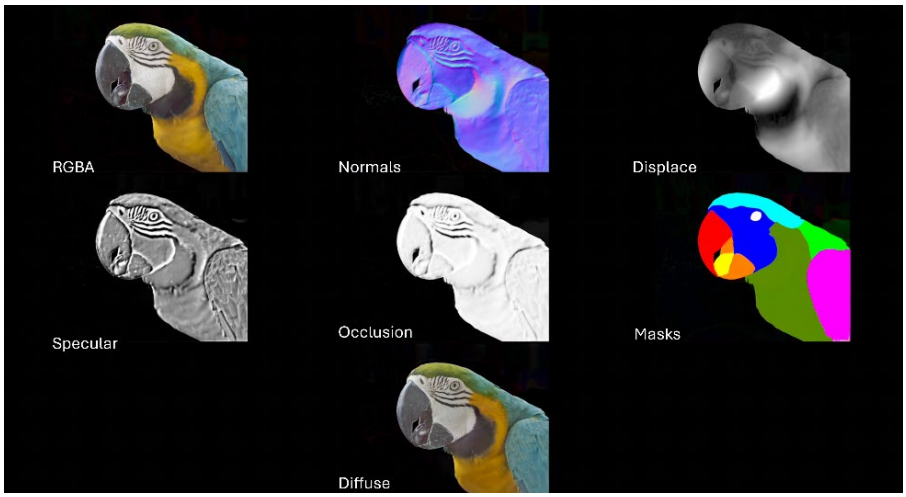


Fig 18. Extracción de una interpretación de diferentes pases 2D.

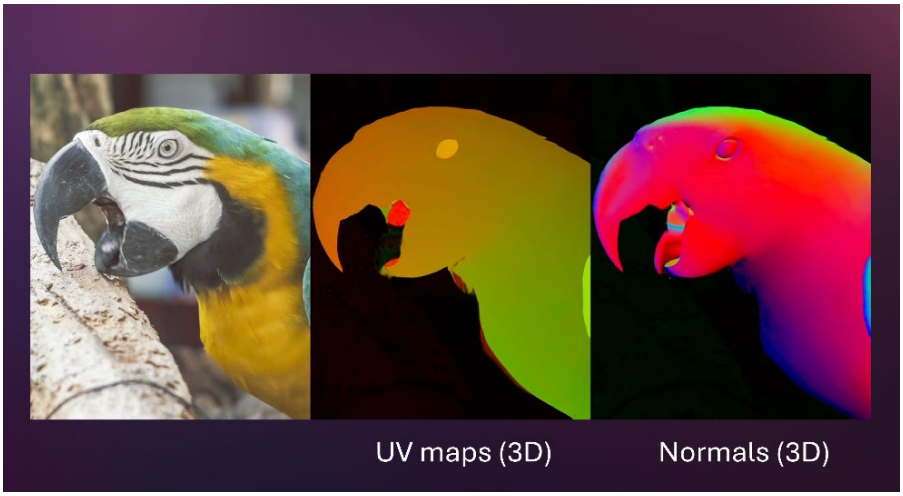


Fig 19. Gracias a Copycat, si contamos con un modelo 3D similar al objeto de la imagen real, podemos recrear pases con información 3D.

Fig 20. Gracias a la información de UVs obtenida mediante el proceso anterior, podemos aplicar nuevas texturas que siguen la animación de la imagen real original.

Fig 21. Una vez Copycat ha sido entrenado para generar imágenes realistas a través de máscaras, es posible realizar modificaciones libres de las mismas.

- **Generación de objetos a partir de modelos**

  Una vez entrenado Nuke Copycat, a partir de un modelo poligonal sencillo, es posible generar objetos complejos con un alto grado de detalle. El entrenamiento del modelo implica enseñarle a reconocer patrones y texturas a partir del modelo básico, permitiéndole luego extrapolar esta información para crear versiones más detalladas y realistas del objeto original.

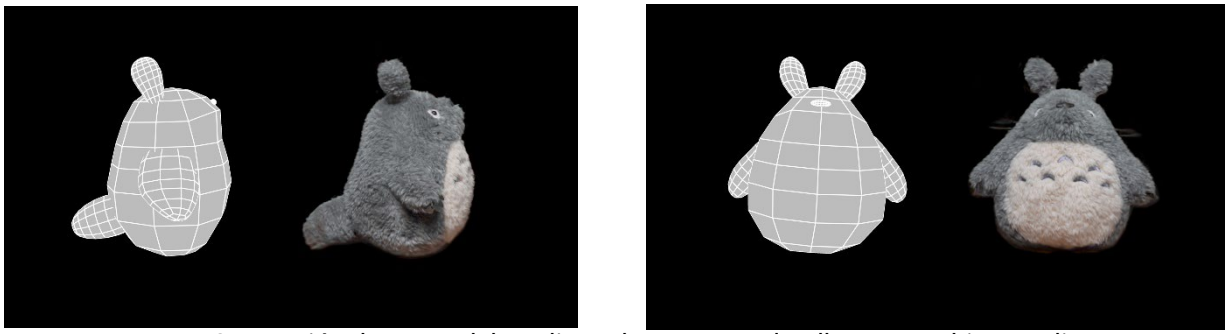


Fig 22. Conversión de un modelo poligonal con pocos detalles en un objeto realista.

Limitaciones

A pesar de su capacidad avanzada para automatizar tareas complejas, Nuke Copycat tiene algunas limitaciones inherentes. El modelo requiere una cantidad significativa de datos de entrenamiento bien etiquetados para alcanzar una alta precisión, lo que puede ser laborioso y consumir mucho tiempo. Además, su rendimiento puede verse comprometido en situaciones donde los patrones o características a reconocer son muy sutiles o altamente variables, lo que podría resultar en resultados inconsistentes. Asimismo, la calidad de los resultados generados está estrechamente vinculada a la calidad del entrenamiento, lo que significa que modelos mal entrenados pueden producir resultados poco fiables.

Fig 23. Ejemplo de una de las limitaciones de Nuke Copycat, en el cual se intenta hacer una equivalencia entre un cilindro 3D y una lata de Coca Cola. Al ser la lata de un material reflectante, el sistema de IA crea artefactos y formas no definidas.

Titularidad o Patentes

En términos de titularidad, patentes y privacidad, CopyCat, al estar integrado en el marco de Nuke, se rige por los mismos principios que rigen a este software desarrollado

por Foundry. El uso de CopyCat está sujeto a las licencias de software de Foundry, que incluyen componentes propietarios y posiblemente patentados.

- **Política de Privacidad:** La política de privacidad de Foundry, aplicable a CopyCat y Nuke, regula cómo se recopilan, usan y protegen los datos personales de los usuarios. Foundry recopila información personal como nombres, direcciones de correo electrónico, y datos de uso del software para mejorar sus productos y servicios. Los datos pueden ser compartidos con terceros bajo circunstancias específicas, como servicios técnicos o requisitos legales. La política también detalla los derechos del usuario sobre sus datos, incluyendo acceso, rectificación y eliminación, conforme a las leyes de protección de datos como el GDPR. Para más información se puede consultar su Política de Privacidad

- **Acuerdo de Licencia:** El acuerdo de licencia de usuario final (EULA) de Foundry, aplicable a CopyCat y Nuke, otorga una licencia limitada, no transferible y no exclusiva para el uso del software. Los usuarios deben cumplir con las restricciones sobre la instalación, uso, distribución y modificación del software. Además, se establecen limitaciones de responsabilidad, reglas sobre la recolección de datos y la necesidad de claves de licencia. El acuerdo también define modelos de licencia específicos, como licencias individuales, de equipo y educativas.

Disponibilidad y Costes de Implantación

Nuke, que incluye la herramienta CopyCat, ofrece diferentes opciones de licencias y precios adaptados a las necesidades de los usuarios. Actualmente, Foundry ha implementado un modelo de suscripción anual para toda la familia de productos Nuke, eliminando gradualmente las licencias perpetuas para nuevos usuarios.

- **Nuke**: La suscripción anual cuesta actualmente 2.839€.
- **NukeX**: Incluye herramientas avanzadas como el rastreo 3D y la corrección de imagen, con un costo actual anual de 4.309€.
- **Nuke Studio**: Ofrece una solución completa de composición y edición por 5.249€ al año.
- **Nuke Indie**: Una versión más asequible diseñada para artistas independientes, con un precio de 449€ anuales.

Para utilizar Nuke de manera eficiente, se recomienda cumplir con los siguientes requisitos de hardware:

- **Sistema Operativo**: Compatible con Windows 10/11, macOS Ventura, y Linux (Rocky 9.0 y CentOS 7.6-7.9).
- **Procesador**: x86-64.
- **Memoria RAM**: Mínimo 8 GB, recomendado 16 GB o más.
- **Almacenamiento**: Al menos 5.7 GB de espacio libre.

- **GPU**: Compatible con OpenGL 2.0. Para tareas avanzadas de aprendizaje automático, se recomienda una GPU NVIDIA compatible con CUDA 11.8 o superior.
- **Pantalla**: Resolución mínima de 1920x1080 píxeles.

Nuke aprovecha la aceleración de hardware mediante el uso de GPU para optimizar el rendimiento en tareas como la visualización de gráficos complejos y la ejecución de nodos que requieren un alto poder de procesamiento. Para maximizar esta aceleración, es esencial contar con una GPU que soporte OpenGL 2.0. Además, para tareas de aprendizaje automático y otras operaciones intensivas en gráficos, se recomienda una GPU NVIDIA con soporte para CUDA 11.8 o superior, lo que permite aprovechar las capacidades de aceleración de hardware de forma más efectiva.

Para aprender a utilizar CopyCat, Foundry ofrece una variedad de recursos educativos y de soporte. Estos incluyen la documentación técnica, tutoriales detallados en su sección "Learn", y videos de capacitación. También cuentan con una comunidad activa de usuarios y un centro de soporte técnico que brinda asistencia personalizada. Además, Foundry proporciona guías específicas y estudios de caso que cubren desde la configuración inicial hasta técnicas avanzadas en CopyCat y su integración en flujos de trabajo. Estos recursos se encuentran accesibles en la página oficial de Foundry Learn.

ComfyUI

Introducción

ComfyUI es una tecnología emergente en el ámbito de la inteligencia artificial aplicada al procesamiento de imágenes y la creación de contenido visual. Lanzada al público en 2023, ComfyUI se destaca como una interfaz gráfica diseñada para facilitar la utilización de modelos de IA en la generación de imágenes, haciendo más accesible la manipulación y creación de contenido visual avanzado. Esta tecnología se ha convertido rápidamente en una herramienta popular entre artistas, desarrolladores y entusiastas de la IA, gracias a su enfoque intuitivo y potente.

ComfyUI fue desarrollada por una comunidad activa y apasionada de desarrolladores independientes y expertos en inteligencia artificial tal y como se muestra en comfy.org. Aunque no es respaldada por una sola empresa, este proyecto se ha beneficiado enormemente del apoyo colectivo de la comunidad de IA, quienes han contribuido con recursos, código y mejoras continuas.

Además de ComfyUI, estos desarrolladores a menudo participan en proyectos de código abierto, contribuyendo a la expansión de herramientas de IA accesibles para todos. La financiación del proyecto proviene en gran parte de donaciones de usuarios,

patrocinadores en plataformas como GitHub y la colaboración de la comunidad, lo que garantiza un desarrollo continuo y orientado a las necesidades de los usuarios.

ComfyUI se centra en proporcionar una interfaz gráfica basada en nodos, simple pero poderosa que permite a los usuarios crear y modificar imágenes mediante el uso de modelos de inteligencia artificial. Sus principales funcionalidades incluyen la capacidad de crear flujos de trabajo para la generación de imágenes, aplicar filtros y ajustes de estilo avanzados, y combinar diferentes modelos de IA para producir resultados únicos y altamente personalizados. En la industria de los efectos visuales (VFX), ComfyUI se destaca por su capacidad de integrar múltiples modelos de IA en procesos creativos, permitiendo a los artistas generar contenido visual de alta calidad de manera más eficiente y con mayor control sobre el resultado final.

Descripción

ComfyUI es una interfaz gráfica nodal diseñada para simplificar y potenciar el uso de inteligencia artificial en la creación y manipulación de imágenes. Esta herramienta se ha convertido en un recurso esencial para profesionales y entusiastas de la industria de VFX que buscan aprovechar el poder de la IA generativa de manera eficiente y controlada.

A continuación, se describen las tres características más destacables:

- Flujo de trabajo visual intuitivo.

ComfyUI permite a los usuarios diseñar flujos de trabajo personalizados mediante un sistema de nodos visuales. Esta característica facilita la experimentación y combinación de diferentes modelos de IA, lo que es especialmente útil en la generación de efectos visuales complejos y detallados, ahorrando tiempo y esfuerzo en el proceso creativo.

- Amplio soporte de modelos de IA Generativos.

En la actualidad ComfyUI soporta un amplio número de modelos de IA Generativo que se pueden consultarse en la siguiente link (Lista de características de ComfyUI). Entre ellos está todo el ecosistema de Stable Diffusion XL. Mediante ComfyUI, los profesionales del VFX tienen a su alcance una herramienta intuitiva que les permite aplicar estilos únicos, realizar modificaciones específicas en las imágenes y generar contenido visual que se alinee perfectamente con las necesidades de su proyecto.

- Extensible y actualizado por una comunidad muy activa.

Gracias a su arquitectura nodal ComfyUI, es fácilmente extensible. Esta característica y una comunidad de usuarios y desarrolladores muy activa es la razón por la que las últimas versiones de todos los modelos de IA Generativa open source suelen estar disponibles para crear workflows nodales en apenas semanas después del lanzamiento de los modelos. Sirva como ejemplo el lanzamiento del modelo Flux de Blackforest, el 1 de Agosto de 2024, que estuvo disponible en ComfyUI el 6 de agosto de 2024.

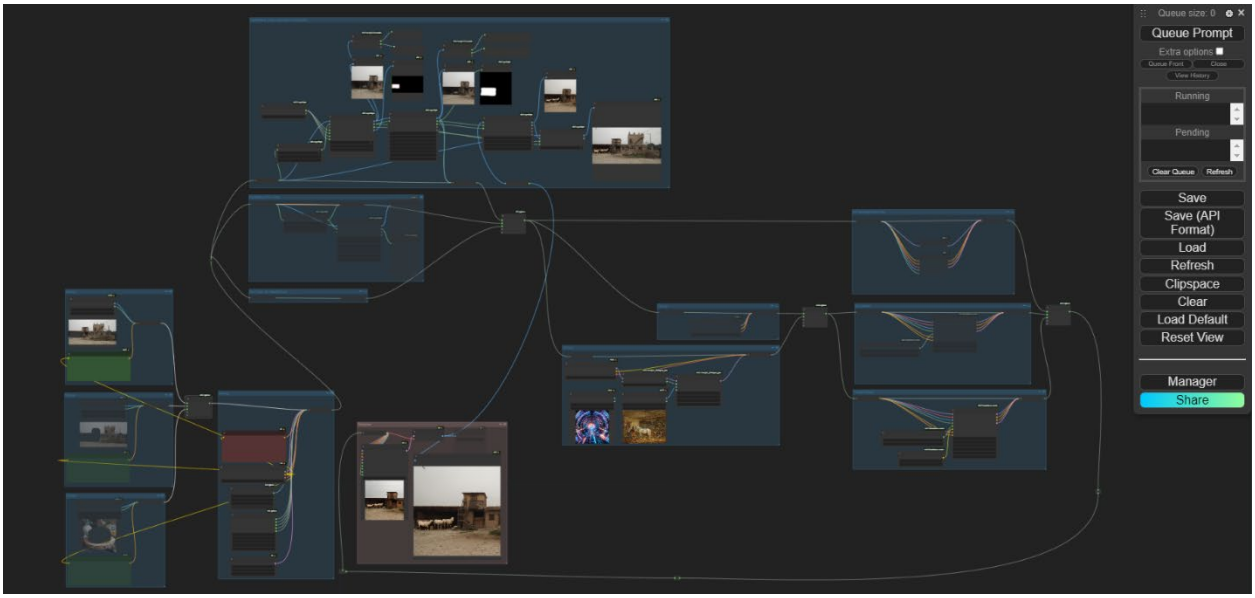


Fig 24. Ejemplo de la interfaz de usuario nodal de ComfyUI

Casos de Uso

ComfyUI es una herramienta diseñada para facilitar la creación y manipulación de imágenes mediante inteligencia artificial. Es ideal para usuarios que buscan una interfaz intuitiva y visual para integrar modelos de IA en sus flujos de trabajo creativos. ComfyUI es especialmente útil en la industria del diseño gráfico, la generación de arte digital y los efectos visuales (VFX). Sin embargo, no es la mejor opción para tareas que requieren un procesamiento masivo de datos o en situaciones donde se necesita una automatización completa sin intervención humana.

- **Casos de uso ideales**

1. Preparación y limpieza de planos de efectos visuales (VFX)

ComfyUI permite a los artistas editar frames de planos de efectos visuales utilizando modelos de IA para crear máscaras de áreas de la imagen y editar dichas áreas a partir de texto e imágenes de referencia, facilitando así está ardua tarea en producciones cinematográficas o publicitarias.

2. Generación de arte digital

Los artistas digitales pueden usar ComfyUI para experimentar con estilos artísticos, combinando diferentes modelos de IA para crear diferentes versiones de forma ágil y controlada.

3. Prototipado rápido de imágenes

Ideal para diseñadores que necesitan iterar rápidamente sobre conceptos visuales, ajustando parámetros de manera intuitiva para obtener resultados inmediatos.

Limitaciones

1. Curva de aprendizaje

Respecto a otras herramientas que simplifican la interfaz de usuario como Automatic111 a costa de su flexibilidad y extensibilidad, ComfyUI tiene una curva de aprendizaje mayor, ya que la creación de grafos de nodos requiere de un mayor conocimiento sobre los diferentes inputs y outputs de los nodos además de conceptos de alto nivel de cómo funcionan los modelos de diffusion como Stable Difussion. A cambio, permite un nivel de personalización y ajuste que es muy necesario para el uso en un entorno profesional de VFX donde cada proyecto introduce problemáticas únicas y específicas.

2. Procesamiento de grandes volúmenes de datos

ComfyUI no es la mejor opción para proyectos que requieren el procesamiento de grandes cantidades de datos o imágenes a escala masiva, donde otras herramientas más especializadas pueden ser necesarias.

3. Requerimientos de hardware

Aunque es una herramienta accesible, su rendimiento óptimo depende de contar con un hardware potente, especialmente en lo que respecta a tarjetas gráficas, lo que puede ser una limitación para algunos usuarios.

Titularidad o Patentes

ComfyUI es una tecnología de código abierto licenciada bajo la GNU General Public License v3.0 (GPL 3.0). Esta licencia garantiza que el software puede ser utilizado, modificado y distribuido libremente, pero con ciertas condiciones que buscan proteger la libertad del software y asegurar que todas las versiones derivadas también se mantengan abiertas.

Licencia de uso

La GPL 3.0 permite a los usuarios utilizar, modificar y distribuir el software, pero requiere que cualquier software derivado o distribuido basado en este código mantenga la misma licencia GPL. Esto significa que cualquier modificación o mejora realizada sobre ComfyUI también debe ser de código abierto si se distribuye. Además, la GPL 3.0 protege contra restricciones que podrían imponerse sobre el software, asegurando que los derechos de los usuarios se mantengan intactos.

Link a la licencia de uso: GPL 3.0 License

Política de privacidad

ComfyUI, al ser un proyecto de código abierto, no recolecta datos de los usuarios de forma directa.

Disponibilidad y Costes de Implantación

ComfyUI es una tecnología de código abierto y se encuentra disponible de forma gratuita bajo la licencia GNU General Public License v3.0 (GPL 3.0). Esto significa que no hay costes directos asociados a su adquisición o uso, lo que la hace accesible para cualquier persona o equipo interesado en aprovechar sus capacidades.

Infraestructura hardware necesaria

Para ejecutar ComfyUI de manera efectiva, se requiere una infraestructura de hardware considerable:

- **Tarjetas Gráficas (GPU):** Se recomienda una GPU moderna con al menos 12 GB de VRAM, como las series NVIDIA RTX 30 o superiores.
- **Memoria RAM:** Mínimo 16 GB, aunque 32 GB o más es ideal para manejar grandes volúmenes de datos.
- **CPU:** Procesadores multi-núcleo con alto rendimiento, como Intel i7/i9 o AMD Ryzen 7/9.
- **Almacenamiento SSD:** Al menos 1TB GB de espacio libre en disco para manejar el modelo y los archivos generados.

Fuentes de ayuda y formación

El equipo que adopte ComfyUI tiene a su disposición una amplia gama de recursos de ayuda y formación:

- **Documentación oficial**: Existe una documentación detallada disponible en línea, que incluye tutoriales, ejemplos de uso y guías de instalación. Esto permite a los usuarios entender rápidamente cómo implementar y personalizar ComfyUI para sus necesidades específicas.
- **Comunidad de usuarios**: ComfyUI cuenta con una comunidad activa de desarrolladores y usuarios que comparten conocimientos a través de foros, canales de Discord y repositorios en GitHub. Esta comunidad es un recurso valioso para resolver dudas, encontrar soluciones a problemas comunes y compartir nuevas ideas o configuraciones.
- **Tutoriales en video y blogs**: Existen numerosos tutoriales en plataformas como YouTube y blogs técnicos que ofrecen guías paso a paso para principiantes, así como contenido más avanzado para usuarios experimentados. Estos recursos permiten al equipo adquirir rápidamente las habilidades necesarias para utilizar ComfyUI de manera efectiva.
- **Cursos y formación**: Aunque no hay cursos oficiales proporcionados por una entidad específica, algunos sitios web educativos y plataformas de aprendizaje en línea pueden ofrecer cursos que cubren tanto el uso de ComfyUI como de las tecnologías relacionadas, como la generación de imágenes con IA y el uso de redes neuronales.

Lenguages y Librerías

Segment Anything

Introducción

El *Segment Anything Model* (SAM) fue desarrollado por el laboratorio Meta AI, y publicado en abril de 2023 a través del artículo académico oficial. SAM es un modelo de segmentación de imágenes diseñado para ser extremadamente versátil y flexible, siendo capaz de identificar y segmentar cualquier objeto dentro de una imágen, sin necesidad de realizar entrenamientos específicos en conjuntos de datos particulares.

La principal motivación detrás de la creación de SAM es la necesidad de herramientas más avanzadas y accesibles para la segmentación de imágenes, una tarea fundamental en muchos campos de aplicación de la Visión Artificial, como puede ser la automatización de la tarea de rotoscopia.

Desde su lanzamiento, SAM ha sido adoptado por una amplia gama de usuarios, incluyendo proyectos de investigación académica, y desarrolladores y empresas tecnológicas que buscar integrar capacidades de segmentación avanzadas en sus productos y servicios. La naturaleza de código abierto de SAM también ha permitido que la comunidad de desarrolladores contribuya a su mejora y expansión, fomentando un ecosistema de colaboración e innovación en torno a esa tecnología.

El lanzamiento del modelo SAM 2, la evolución del *Segment Anything Model*, marca un nuevo hito en el campo de la segmentación de imágenes y video. Se anuncia que SAM 2 pretende mejorar significativamente en términos de precisión y rendimiento, con aplicación directa en vídeos en tiempo real. Además, según los anuncios oficiales, será optimizado para ofrecer tiempos de procesamiento más rápidos, haciendo que su implementación sea aún más eficiente para casos de uso industriales y comerciales. En próximas versiones de este documento, se explorarán en mayor detalle las mejoras técnicas introducidas en SAM 2, así como ejemplos de aplicaciones prácticas que ya están aprovechando esta nueva tecnología.

Fig 25. Máscaras generadas con SAM

Descripción

SAM es un modelo poderoso y flexible diseñado para la segmentación precisa y eficiente de elementos en imágenes, permitiendo crear máscaras de cualquier objeto mediante el uso de técnicas avanzadas de Visión Artificial. El modelo ha sido entrenado para reconocer y delinear objetos sin necesidad de entrenamiento adicional específico, lo que lo hace extremadamente versátil y fácil de usar en una variedad de aplicaciones. Entre las características técnicas principales se encuentran:

- **Segmentación Universal.** SAM está diseñado para segmentar cualquier tipo de objeto en una imagen, sin necesidad de datos de entrenamiento específicos para cada tipo de objeto. Esto se logra mediante un modelo pre entrenado en grandes volúmenes de datos variados.

- **Multimodalidad.** Este modelo puede manejar diferentes tipos de datos de entrada, como imágenes estáticas, secuencias de video y datos tridimensionales, ampliando su aplicabilidad a diversos campos.

- **Alto Rendimiento y Escalabilidad.** Diseñado para ser eficiente y escalable, SAM puede procesar grandes volúmenes de imágenes rápidamente, lo cual es ideal para aplicaciones que requieren segmentación en tiempo real o en grandes cantidades de datos.

- **Personalización y Ajuste.** Aunque SAM es potente en su configuración predeterminada, los usuarios tienen la posibilidad de personalizar y ajustar el modelo según sus necesidades específicas, optimizando su rendimiento para tareas particulares.

- **Soporte y Actualizaciones Continuas.** Al ser de código abierto, SAM se beneficia de una comunidad activa de desarrolladores y usuarios que contribuyen a su mejora continua, así como de actualizaciones regulares que incorporan las últimas investigaciones y avances en segmentación de imágenes.

Casos de Uso

SAM es el modelo más avanzado en tareas de segmentación por IA, sin embargo, y al igual que cualquier herramienta tecnológica presenta una serie de casos de uso óptimos, pero también ciertas limitaciones:

Casos de uso ideales

- Pre Segmentación. Utilizar SAM para pre segmentar escenas permite a los artistas enfocarse en la refinación y el ajuste final, en lugar de empezar desde cero.

- Consistencia en la segmentación. Al usar un modelo pre entrenado, SAM asegura una consistencia alta en la segmentación de objetos a través de diferentes fotogramas, lo cual es crucial en la rotoscopia.
- Trabajo en grandes volúmenes de datos. SAM puede manejar y procesar grandes volúmenes de imágenes de manera eficiente, permitiendo la segmentación de secuencias de vídeo compuestas por un gran número de fotogramas.
- Detección precisa. SAM es eficaz en la detección de bordes y detalles finos, mejorando la calidad de la rotoscopia en escenas complejas.
- Integración con otros software y herramientas. La capacidad de SAM para integrarse con otros software y herramientas de edición facilita su uso en flujos de trabajo existentes.

Limitaciones

- Movimiento Rápido y Complejo. En escenas donde los objetos tienen movimientos extremadamente rápidos o complejos, SAM puede tener dificultades para mantener una segmentación precisa en todos los fotogramas.
- Cambios de Forma y Apariencia. Si los objetos cambian drásticamente de forma o apariencia entre fotogramas, SAM puede necesitar ajustes significativos y no ser tan eficiente.
- Iluminación Inconsistente. En escenas donde la iluminación cambia drásticamente de un fotograma a otro, SAM puede tener problemas para segmentar correctamente, ya que la variabilidad de la iluminación puede afectar su precisión.
- Recursos Computacionales Insuficientes. SAM requiere un hardware potente para funcionar de manera eficiente. Si el entorno de trabajo no cuenta con los recursos computacionales necesarios, el rendimiento y la velocidad de procesamiento pueden ser inadecuados para las necesidades de producción.

Titularidad o Patentes

El Segment Anything Model (SAM) es una tecnología desarrollada por Meta AI y está disponible como una librería de código abierto. Esto significa que no es una tecnología propietaria en el sentido estricto, ya que su código fuente está disponible para ser utilizado, modificado y distribuido por cualquiera bajo los términos de su licencia.

SAM está licenciado bajo la licencia Apache 2.0, una licencia de **software libre** permisiva que permite a los usuarios utilizar el software para cualquier propósito, distribuirlo, modificarlo y distribuir versiones modificadas del software bajo los mismos términos. La licencia Apache 2.0 también proporciona una cláusula explícita de patente,

que protege a los usuarios del software de ser demandados por patentes aplicables al software.

Esta política de privacidad describe cómo se manejan y protegen los datos al utilizar el Segment Anything Model (SAM) desarrollado por Meta AI cuando se implementa y ejecuta **localmente en un entorno propio**. Meta AI no recopila, almacena ni tiene acceso a ningún dato procesado a través de SAM cuando se utiliza en local.

Datos Recopilados

AI utilizar SAM en un entorno local todas las imágenes y datos relacionados que se procesan con SAM se mantienen exclusivamente en el sistema local, Meta AI no recopila, almacena ni accede a estos datos. No se recopilan datos de uso del modelo cuando se ejecuta localmente, ya que la operación y el procesamiento ocurre enteramente en su entorno privado.

Uso de los Datos

Los datos de las imágenes se utilizan únicamente para los fines de segmentación según sus necesidades y se mantienen en su control exclusivo. Todos los datos procesados permanecen confidenciales y no son accesibles para Meta AI ni para terceros.

Almacenamiento de Datos

Todos los datos se almacenan en el hardware y sistemas locales, a los cuales Meta AI no tiene acceso. El desarrollador es responsable de implementar las medidas de seguridad adecuadas para proteger los datos.

Disponibilidad y Costes de Implantación

Hardware Recomendado

Para desarrollar y ejecutar el Segment Anything Model (SAM) de manera eficiente, es recomendable contar con una infraestructura de hardware adecuada que pueda manejar las demandas computacionales del modelo. A continuación se describe el hardware necesario y recomendado:

- **GPU.** Para obtener un rendimiento óptimo, es esencial utilizar una tarjeta gráfica de alto rendimiento compatible con CUDA. Se recomienda especialmente el uso de tarjetas gráficas de las series NVIDIA RTX, conocidas por su capacidad para manejar tareas intensivas de procesamiento gráfico y aprendizaje profundo. Alternativamente, para aplicaciones de gama profesional, las tarjetas NVIDIA A100 ofrecen un

rendimiento superior, optimizado para cargas de trabajo de inteligencia artificial y procesamiento de datos a gran escala.

- **Memoria RAM.** Es aconsejable disponer de al menos 32 GB de RAM para asegurar que el sistema pueda manejar grandes volúmenes de datos y realizar procesamiento en paralelo sin interrupciones. Una memoria RAM adecuada es crucial para soportar las operaciones intensivas del modelo SAM, permitiendo un flujo de trabajo más fluido y eficiente.

- **Sistema de Refrigeración.** Un sistema de refrigeración eficiente es crucial para mantener el hardware a temperaturas operativas seguras, especialmente durante el procesamiento intensivo de imágenes. La implementación de soluciones de refrigeración líquida o sistemas avanzados de ventilación puede prevenir el sobrecalentamiento, garantizando que los componentes de hardware funcionen de manera óptima y prolongando su vida útil.

Contar con este conjunto de hardware recomendado no solo mejora la eficiencia y la velocidad de procesamiento del Segment Anything Model, sino que también asegura la estabilidad y fiabilidad del sistema en tareas computacionalmente intensivas. Esta configuración de hardware es ideal para maximizar el rendimiento y obtener resultados precisos y rápidos en aplicaciones de segmentación de imágenes.

Fuentes de Ayuda y Formación para un Desarrollo con SAM

Para que el equipo pueda desarrollar con éxito utilizando el *Segment Anything Model* (SAM), es fundamental contar con diversas fuentes de ayuda y formación. A continuación, se describen algunas de las principales fuentes disponibles:

- **Documentación Oficial**
    1. Repositorio de GitHub: El repositorio oficial de SAM en GitHub es una fuente esencial de información. Este repositorio incluye la documentación completa del modelo, guías detalladas de instalación, ejemplos de código práctico y detalles exhaustivos sobre las funcionalidades del modelo. El acceso al repositorio proporciona a los desarrolladores todas las herramientas necesarias para comenzar y avanzar en el uso de SAM.

    2. Wiki y FAQs: Dentro del repositorio de GitHub, se encuentra una sección de Wiki o Preguntas Frecuentes (FAQs) que ofrece respuestas a las preguntas más comunes y soluciones a problemas frecuentes que los desarrolladores pueden encontrar. Además, esta sección proporciona

guías de buenas prácticas, lo que ayuda a optimizar el uso del modelo y a evitar errores comunes.

- **Tutoriales, ejemplos y notebooks interactivos**

  La documentación oficial suele incluir tutoriales paso a paso que cubren desde la instalación básica del modelo hasta ejemplos avanzados de su uso en diversos contextos. Estos tutoriales están diseñados para facilitar la comprensión del modelo y su aplicación práctica. Además, se proporcionan notebooks interactivos, que permiten a los usuarios interactuar directamente con el modelo en un entorno práctico y dinámico. Estos notebooks son especialmente útiles para aprender sobre el funcionamiento interno de SAM y experimentar con diferentes configuraciones y datos.

Contar con estas fuentes de ayuda y formación es crucial para asegurar un desarrollo eficiente y exitoso con el *Segment Anything Model*. Estas herramientas no solo proporcionan la información necesaria para la instalación y uso del modelo, sino que también ofrecen un soporte continuo para resolver dudas y optimizar el proceso de desarrollo.

YOLOv8

Introducción

YOLO (*You Only Look Once*) es una familia de modelos para la detección de objetos en imágenes que se ha desarrollado y evolucionado en varias versiones a lo largo de los últimos años. Inicialmente, esta librería fue creada por Joseph Redmon entre otros investigadores de la Universidad de Washington en 2016, y su versión original se presentó en el artículo académico de investigación *"You Only Look Once: Unified, Real-Time Object Detection"*.

La versión más reciente, YOLOv8, ha sido desarrollada por Ultralytics, una empresa dedicada al desarrollo de soluciones y modelos de IA, concretamente de Visión Artificial. Esta empresa es una de las fuerzas principales detrás, no solo de esta versión actual, sino de muchas de las anteriores evoluciones del modelo. La versión v8 de YOLO presenta avances notables en términos de rendimiento, velocidad y precisión, siendo también adaptable fácilmente a distintas plataformas y aplicaciones.

Esta librería es utilizada por una amplia variedad de usuarios y sectores de aplicación, debido a sus capacidades avanzadas de detección en tiempo real. Sus aplicaciones varían desde la investigación y el entorno académico, hasta el mundo empresarial y sectores industriales con aplicaciones como la vigilancia, la automatización y soluciones avanzadas de reconocimiento de objetos, entre otras.

La popularidad de YOLOv8 se debe a su capacidad para realizar detecciones rápidas y precisas, siendo ideal para aplicaciones en tiempo real, constituyendo la librería como una herramienta clave en el ámbito de la Visión Artificial.

Descripción

La funcionalidad principal de esta librería es identificar y localizar múltiples objetos en recursos visuales, ya sea en una única imagen o en una secuencia de vídeo. Para ello, utiliza una arquitectura de Red Neuronal Convolucional (CNN) optimizada para llevar a cabo estas tareas de forma eficiente, haciendo de YOLOv8 la opción ideal para aplicaciones que requieren de un procesamiento ágil y rápido. Entre sus características principales, cabe destacar:

- **Alta precisión.** La versión más reciente de YOLO mejora la precisión de detección en comparación con versiones anteriores y otros modelos que cumplen la misma función. Utiliza técnicas avanzadas de *Deep Learning* para mejorar la exactitud de las predicciones y reducir la tasa de error en la detección.
- **Flexibilidad y Escalabilidad.** Se trata de un modelo altamente flexible, el cual puede ser entrenado o ajustado para detectar una amplia variedad de objetos. Los usuarios tienen la opción de ajustarlo a sus necesidades específicas mediante el entrenamiento con conjuntos de datos personalizados.
- **Optimización de Recursos.** A pesar de su alta capacidad de procesamiento, YOLOv8 está optimizado para funcionar eficientemente en hardware diverso, desde potentes GPUs hasta dispositivos más limitados. Los requerimientos de hardware vienen determinados por la tarea a realizar, ya que el modelo es altamente adaptable.
- **Compatibilidad y Soporte.** Esta librería es compatible con la gran mayoría de *frameworks* populares para desarrollos de IA, incluyendo PyTorch y TensorFlow, lo cual facilita su integración en diversas plataformas y desarrollos.

En el contexto de un desarrollo para Segmentación Asistida por IA, el uso de un modelo con las capacidades y características de YOLOv8 permite hacer un seguimiento del elemento a segmentar a lo largo de todos los fotogramas, de una forma eficiente y precisa. La "caja de recorte" o *"bounding-box"* que contiene el objeto en cada imagen sirve también como entrada para el modelo de segmentación, limitando el margen de error con respecto a otros elementos de la imagen y automatizando el proceso de *tracking*.

Fig 26. Imágenes publicadas por los desarrolladores de YOLOv8

Casos de Uso

- **Casos de uso ideales**

El Motion Tracking asistido por IA es especialmente efectivo en una variedad de escenarios donde sus capacidades pueden aprovecharse al máximo. A continuación, se enumeran algunos de los casos de uso ideales para esta tecnología:

- **Escenas con fondos claros**. YOLOv8 puede ser muy efectivo para la detección y seguimiento de objetos en escenas con fondos uniformes o de bajo detalle, donde los bordes de los objetos están claramente definidos. En estos escenarios, la claridad del fondo permite que los algoritmos de detección identifiquen y rastreen los objetos con mayor precisión y eficiencia.

- **Objetos en movimiento**. La detección de objetos que se mueven de manera predecible y no se superponen significativamente con otros objetos puede ser gestionada de manera eficiente por YOLOv8. Este caso de uso es ideal para aplicaciones donde los objetos tienen trayectorias claras y no experimentan oclusiones frecuentes, facilitando un seguimiento continuo y preciso.

- **Delimitar el objeto o personaje a segmentar.** YOLOv8 coloca una caja de recorte alrededor del objeto a identificar y trackear, lo que permite una

mejor segmentación del personaje u objeto al delimitar correctamente la zona de la imagen en la que se ubica. Esta funcionalidad es crucial para asegurar que el área de interés se procese con mayor precisión, mejorando la calidad de la segmentación y el seguimiento.

- **Limitaciones**

El Motion Tracking asistido por IA puede enfrentar desafíos significativos en ciertos escenarios complejos. A continuación, se describen algunos casos de uso menos adecuados para esta tecnología:

- **Escenas con Múltiples Objetos Superpuestos**. El sistema experimenta ciertas dificultades en escenas con muchas personas u objetos en movimiento, donde los elementos se superponen frecuentemente. Esto puede disminuir el seguimiento preciso de cada individuo o elemento, ya que los algoritmos pueden confundir los objetos o perder el rastro de ellos.

- **Entornos complejos y ricos en detalle.**Escenarios con muchos detalles, como bosques densos, ciudades con mucho tráfico o interiores llenos de objetos, presentan un reto considerable. La gran cantidad de elementos visuales similares y el alto nivel de detalle pueden confundir al algoritmo de seguimiento, reduciendo la precisión del motion tracking.

- I**nteracciones complejas y dinámicas.** Escenas de lucha o acción intensa, donde los personajes interactúan de manera muy dinámica y cambiante con su entorno, son especialmente difíciles para el motion tracking asistido por IA. Los movimientos rápidos y erráticos, junto con las interacciones constantes entre los personajes y el entorno, pueden superar las capacidades de los algoritmos actuales, resultando en un seguimiento impreciso.

Titularidad o Patentes

YOLOv8, desarrollado por Ultralytics, está disponible bajo dos tipos principales de **licencias**:

1. Licencia AGPL-3.0: Esta es una licencia de código abierto aprobada por la Open Source Initiative (OSI). Está diseñada para promover la colaboración y el intercambio de conocimientos. Bajo esta licencia, cualquier software que incorpore código o modelos de Ultralytics debe ser también de código abierto. Esta opción es ideal para proyectos de investigación y proyectos no comerciales .

2. **Licencia Empresarial**: Esta licencia está destinada a usos comerciales. Permite integrar los modelos y el software de Ultralytics en productos y servicios comerciales sin la obligación de abrir el código fuente. Es adecuada para empresas que desean utilizar las capacidades de YOLOv8 en productos propietarios. Las empresas pueden solicitar esta licencia a través de un formulario en el sitio web de Ultralytics.

Con respecto a las **políticas de privacidad**, Ultralytics se compromete a proteger la privacidad de los usuarios de su software. La empresa recopila datos de uso de forma anónima para mejorar la experiencia del usuario y la funcionalidad de sus paquetes de Python, incluyendo los modelos avanzados de YOLO. Esta recopilación de datos incluye métricas de uso, información del sistema y datos de rendimiento, los cuales son procesados de manera agregada y anónima para asegurar la privacidad del usuario.

Además, Ultralytics utiliza Google Analytics para analizar el tráfico y el uso del software, y Sentry para el reporte de errores y crashes, siempre asegurando que la información recopilada no contenga datos identificables personalmente. Los usuarios tienen la opción de desactivar la recopilación de datos modificando la configuración del paquete.

Disponibilidad y Costes de Implantación

- **Licencias y Precios**

La única licencia que requiere de un pago por parte del usuario es la licencia empresarial, el coste de esta licencia no es fijo y puede variar según el alcance y la escala del uso previsto. Los detalles específicos y las cotizaciones personalizadas deben ser solicitados directamente a través del formulario de contacto en el sitio web de Ultralytics.

- **Infraestructuras hardware recomendadas**

La ejecución de YOLOv8, al igual que muchos otros modelos de Visión Artificial, se basa en el uso de Redes Neuronales Profundas y requiere de cierta infraestructura de hardware para garantizar un rendimiento óptimo, especialmente en términos de velocidad y precisión. A continuación, se presentan las recomendaciones de configuraciones de hardware para distintos escenarios, detallando las especificaciones necesarias para desarrollo, pruebas locales y despliegue en producción.

- **Desarrollo y Pruebas Locales**

Para llevar a cabo el desarrollo y las pruebas locales de modelos de Visión Artificial como YOLOv8, se recomienda la siguiente configuración de hardware:

- **CPU:** Procesadores con múltiples núcleos y alta frecuencia de reloj que permiten un procesamiento eficiente de tareas complejas y paralelizadas, como es el caso de los equipos AMD Ryzen 5 o superiores
- **GPU:**Tarjetas gráficas con una arquitectura optimizada para el procesamiento de tareas de aprendizaje profundo, proporcionando una aceleración significativa en la ejecución de modelos de redes neuronales, como puede ser el caso de Nvidia RTX 3060 o superiores
- **RAM:** Se recomienda un mínimo de 32 GB, ya que una cantidad adecuada de memoria RAM es esencial para manejar grandes volúmenes de datos y operaciones en memoria, asegurando un rendimiento fluido durante el entrenamiento y la inferencia de modelos.

- **Despliegue en Producción**

Para el despliegue en entornos de producción, donde se espera un rendimiento constante y robusto, se recomienda una configuración de hardware más avanzada:

- **CPU:** Procesadores de alta gama con mayor número de núcleos y capacidades de procesamiento, adecuados para manejar cargas de trabajo intensivas y sostenidas, como AMD Ryzen 9 o superiores.
- **GPU:** Tarjetas gráficas de nivel profesional con una gran cantidad de memoria VRAM, diseñadas para aplicaciones de IA que requieren un procesamiento intensivo de gráficos y datos, proporcionando una alta eficiencia y rendimiento en tareas de inferencia y entrenamiento de modelos a gran escala. NVIDIA RTX A6000, NVIDIA Quadro RTX 8000 (ambas con 48 GB de VRAM), o superiores
- **RAM:** Mínimo 32 GB, siendo óptimo a partir de 48 GB. Una mayor cantidad de memoria RAM es crucial en entornos de producción para asegurar que los modelos puedan operar sin restricciones de memoria, permitiendo un procesamiento rápido y eficiente de grandes conjuntos de datos y operaciones concurrentes.

Fuentes de Ayuda y Formación para el uso de YOLOv8 en Desarrollo

Para el uso de YOLOv8 en un desarrollo propio se dispone de diversas fuentes de ayuda y formación que permitirán aprovechar al máximo esta herramienta avanzada de detección de objetos. Estas fuentes de recursos son fundamentales para asegurar una

implementación eficiente y efectiva de la tecnología. A continuación, se describen las principales fuentes de recursos disponibles:

- **Documentación Ultralytics.** Ultralytics proporciona una documentación completa y detallada para su producto YOLOv8. Esta documentación abarca todos los aspectos necesarios para utilizar la herramienta de manera efectiva, incluyendo guías exhaustivas de instalación, configuración, entrenamiento de modelos y despliegue en producción. Las guías de instalación proporcionan instrucciones paso a paso para la correcta configuración del entorno y la instalación del software, asegurando que los usuarios puedan empezar a utilizar YOLOv8 sin problemas técnicos.

  Además, la documentación incluye secciones dedicadas a la configuración, permitiendo a los usuarios personalizar la herramienta según sus necesidades específicas. Las guías de entrenamiento de modelos ofrecen una descripción detallada del proceso de entrenamiento, ayudando a los usuarios a crear modelos precisos y eficientes. Por último, la documentación también cubre el despliegue en producción, proporcionando las mejores prácticas y soluciones para integrar YOLOv8 en entornos de producción de manera efectiva.

- **Discusiones e Issues de Github.** El repositorio de YOLOv8 en GitHub es una fuente valiosa de información donde los usuarios pueden plantear preguntas, reportar problemas y compartir soluciones. Este espacio de colaboración y comunidad es fundamental para el desarrollo y la mejora continua de la herramienta.

  - **Discusiones**

    En la sección de discusiones del repositorio, los usuarios pueden iniciar conversaciones sobre diversos temas relacionados con YOLOv8. Estas discusiones permiten el intercambio de conocimientos, experiencias y mejores prácticas entre usuarios de diferentes niveles de experiencia. Los temas pueden variar desde la configuración y uso básico de la herramienta hasta la implementación de características avanzadas y la optimización del rendimiento.

  - **Issues**

    La sección de issues (problemas) del repositorio de GitHub es un lugar donde los usuarios pueden reportar errores, solicitar nuevas funcionalidades y hacer preguntas técnicas. Cuando un usuario encuentra un problema o un bug, puede crear un issue detallando el problema, proporcionando ejemplos y, si es posible, sugerencias para su solución. Los desarrolladores y otros miembros de la comunidad pueden entonces investigar, discutir y trabajar juntos para resolver estos problemas.

- **Soporte de Ultralytics.** Ultralytics ofrece un soporte técnico robusto y directo a través de su plataforma de soporte, conocida como Ultralytics Hub. Este soporte es especialmente valioso para usuarios empresariales que requieren asistencia técnica avanzada para implementar y optimizar las soluciones de IA en sus proyectos. A través de Ultralytics Hub, los usuarios pueden acceder a una variedad de recursos, incluyendo documentación detallada, tutoriales y la posibilidad de interactuar directamente con el equipo de soporte técnico de Ultralytics. Este servicio está disponible mediante sus planes de soporte, que están diseñados para ofrecer diferentes niveles de asistencia según las necesidades específicas de cada cliente, asegurando que las empresas puedan maximizar el valor y la eficiencia de las herramientas de Ultralytics en sus aplicaciones.

OpenCV. Optical Flow. RLOF

Introducción

El modelo RLOF (*Robust Local Optical Flow*) es una técnica avanzada dentro de la librería OpenCV, diseñada para calcular el flujo óptico de manera robusta y eficiente. OpenCV, que significa *Open Source Computer Vision Library*, es una biblioteca de visión por computadora de código abierto creada inicialmente por Intel en el año 1999 y posteriormente apoyada por Willow Garage y Itseez (adquirida de nuevo por Intel en 2016).

OpenCV fue creado con el objetivo de proporcionar una infraestructura común para aplicaciones de visión por computadora y para acelerar el uso de la percepción basada en el aprendizaje en productos comerciales. Con el paso del tiempo, OpenCV se ha convertido en una de las bibliotecas más populares en el campo de la visión por computadora, siendo ampliamente utilizada tanto en la industria como en la academia.

El algoritmo RLOF fue incorporado en OpenCV para ofrecer una solución más robusta al problema del flujo óptico, el cual es esencial para tareas como el seguimiento de objetos, la estabilización de video y la estimación del movimiento. Este modelo se destaca por su capacidad para manejar grandes variaciones de iluminación y condiciones de movimiento, lo que lo hace particularmente útil en aplicaciones del mundo real donde las condiciones pueden ser adversas o cambiantes.

El modelo RLOF es utilizado por una amplia gama de profesionales y organizaciones. Entre estos se encuentra el ámbito de investigación académica donde se emplea para desarrollar nuevos algoritmos y modelos de visión artificial, y también en empresas y el sector de desarrollo que lo integran en sus productos para mejorar funcionalidades de *tracking* o realidad aumentada entre otras.

Descripción

El modelo RLOF (Robust Local Optical Flow) en OpenCV está diseñado para calcular el flujo óptico, es decir, el movimiento aparente de los objetos, superficies y bordes en una secuencia de imágenes. Esta técnica es crucial en aplicaciones que requieren seguimiento de movimiento y análisis de vídeo, como es el caso de *tracking* para la asistencia en tareas de rotoscopia. RLOF se basa en la estimación local del flujo óptico, lo que significa que analiza pequeñas regiones de la imagen para determinar el desplazamiento de píxeles entre fotogramas consecutivos. Entre sus características principales destacan las siguientes:

- **Robustez.** RLOF es notable por su capacidad para seguir objetos y detectar movimiento incluso en condiciones adversas, como cambios significativos en la iluminación de la escena. Este aspecto es particularmente desafiante para muchos algoritmos de flujo óptico, que suelen depender de condiciones de iluminación constantes para mantener la precisión. La habilidad de RLOF para adaptarse a variaciones en la luz le permite ofrecer un rendimiento superior en una amplia gama de escenarios, desde ambientes exteriores con luz natural cambiante hasta interiores con luces artificiales variables.

  Además, RLOF muestra una destacable robustez en entornos ruidosos. El ruido en las imágenes, que puede surgir de diversas fuentes como, condiciones meteorológicas adversas o interferencias electrónicas, tiende a afectar negativamente la precisión de la detección de movimiento en muchos modelos de flujo óptico. Sin embargo, RLOF ha sido diseñado para mantener su precisión incluso en presencia de ruido significativo. Esta capacidad le permite operar eficazmente en condiciones donde otros algoritmos podrían fallar, garantizando una detección del movimiento confiable y consistente.

- **Precisión.** El uso de RLOF (Robust Local Optical Flow) se destaca por su alta precisión, especialmente cuando se enfoca en regiones pequeñas dentro de una imagen o secuencia de vídeo. Esta capacidad para detectar movimientos sutiles con gran exactitud es fundamental para aplicaciones que requieren un seguimiento fino y detallado, como en la animación y los efectos visuales avanzados y estudios de biomecánica. La precisión de RLOF se debe a su metodología robusta, que minimiza los errores y permite un análisis detallado de los movimientos más leves.

  Además, RLOF ofrece la posibilidad de ajustar una variedad de parámetros, lo que permite optimizar la estimación del flujo óptico según las necesidades específicas de cada aplicación. Esta flexibilidad es crucial, ya que diferentes proyectos pueden requerir configuraciones personalizadas para obtener los mejores resultados. Por ejemplo, en situaciones donde los movimientos son extremadamente pequeños o cuando se trabaja con fondos complejos, la capacidad de ajustar estos parámetros garantiza que el

seguimiento sea preciso y fiable. Esta característica hace que RLOF sea una herramienta versátil y adaptable a una amplia gama de contextos y exigencias técnicas.

- **Eficiencia Computacional.** Este software stá optimizado para funcionar en tiempo real, lo que lo convierte en una herramienta eficaz para aplicaciones que requieren una rápida respuesta y procesamiento instantáneo de datos visuales. Esta optimización permite que RLOF sea utilizado en una amplia gama de escenarios donde la velocidad es crítica. Además, RLOF es compatible con diversas arquitecturas de hardware, lo que significa que puede ser implementado en una variedad de dispositivos, incluyendo aquellos con recursos computacionales limitados. Esta compatibilidad amplia asegura que RLOF puede funcionar eficientemente en dispositivos móviles, sistemas embebidos y otros equipos que no disponen de capacidades de procesamiento robustas. Esta versatilidad es particularmente beneficiosa en contextos donde la implementación en hardware económico y de bajo consumo es una prioridad, permitiendo a los desarrolladores utilizar RLOF en aplicaciones más accesibles y variadas sin comprometer el rendimiento.

En el contexto del uso del modelo en técnicas de segmentación y tracking, el uso de RLOF es clave para complementar la tarea de *tracking* de YOLO pero en este caso aplicado a los puntos de control del usuario, no al objeto a segmentar. Esto permite que el usuario solo tenga que proporcionar el input en el primer fotograma y gracias a estos algoritmos, se calculan todas las posiciones relativas de los puntos de control y el objeto en todos los demás fotogramas.

Casos de Uso

Casos de uso ideales

A continuación, se detallan algunos de los casos de uso ideales para RLOF, donde su implementación puede ofrecer ventajas significativas en términos de rendimiento y eficiencia.

- **Seguimiento de movimientos no bruscos.** El seguimiento de objetos en movimiento es una capacidad fundamental en la rotoscopia asistida por IA, y la herramienta RLOF (Robust Local Optical Flow) destaca en esta área por su eficacia y precisión. RLOF es particularmente ideal para seguir movimientos suaves en la rotoscopia, permitiendo la creación de máscaras precisas para objetos que se desplazan de manera uniforme. Ejemplos de

tales movimientos incluyen una persona caminando o un vehículo en movimiento regular. La tecnología de flujo óptico local que emplea RLOF facilita la captura detallada de estos movimientos suaves, garantizando que los contornos de los objetos se mantienen nítidos y coherentes a lo largo de la secuencia de vídeo.

No obstante, RLOF no se limita únicamente a escenas con movimientos uniformes. Esta herramienta es notablemente robusta y versátil, lo que le permite realizar un seguimiento preciso incluso en escenas con movimientos más complejos y cambios imprevistos. Por ejemplo, en situaciones donde los objetos cambian de dirección abruptamente o donde la velocidad de movimiento varía considerablemente, RLOF puede ajustarse y mantener la precisión del seguimiento.

- **Estabilización de videos.** En el ámbito de la postproducción, la técnica de Optical Flow basada en Redes se puede utilizar para la estabilización de videos, desempeñando un papel crucial en la mejora de la calidad visual. Esta técnica permite eliminar vibraciones y movimientos bruscos que pueden ocurrir durante la grabación. Al aplicar RLOF, se consigue una imagen final más fluida y estable. Esta estabilización mejora la y facilita significativamente el trabajo de los sistemas de segmentación automatizada, como Segment Anything. Al presentar un video más uniforme los algoritmos de segmentación pueden identificar y delinear objetos con mayor precisión y consistencia, mejorando la eficiencia y precisión del proceso de segmentación.

- **Análisis de detalles finos del movimiento**. El análisis de detalles finos del movimiento es una capacidad crucial para aplicaciones que requieren una rotoscopia precisa, especialmente en escenas complejas. La técnica RLOF puede ser de gran ayuda en este aspecto ya que es capaz de capturar y seguir detalles minuciosos del movimiento, lo cual es esencial para obtener una rotoscopia precisa.La capacidad de capturar estos detalles finos es fundamental no solo para la precisión visual, sino también para mantener la coherencia y realismo de las máscaras en la escena. Sin esta capacidad, los movimientos sutiles del cabello o la ropa podrían parecer artificiales o incorrectos, rompiendo la inmersión y calidad visual de la escena.

Limitaciones

Aunque la técnica de Optical Flow ofrece ventajas significativas en la captura de detalles finos del movimiento, no está exenta de limitaciones. Es importante reconocer estas limitaciones para entender mejor en qué contextos esta tecnología puede no ser la más adecuada y dónde podrían ser necesarios ajustes adicionales o el uso de

técnicas complementarias. A continuación, se explorarán las principales limitaciones de RLOF en la aplicación de motion tracking y rotoscopia asistida por IA.

- **Oclusiones frecuentes.** El algoritmo RLOF puede enfrentar serias dificultades para mantener un seguimiento preciso en presencia de oclusiones. Las oclusiones ocurren cuando el objeto de interés es temporalmente bloqueado por otros objetos o por partes del entorno. En estos casos, el algoritmo puede perder la pista del objeto bloqueado, ya que su capacidad para predecir y seguir la trayectoria del objeto se ve comprometida por la falta de visibilidad continua. Esta limitación puede resultar en un seguimiento interrumpido y en la necesidad de intervención manual para corregir la trayectoria del objeto una vez que vuelve a ser visible.

- **Texturas uniformes o con bajo contraste.** El modelo RLOF requiere puntos característicos bien definidos en el objeto a rastrear para funcionar correctamente. En superficies uniformes, donde la distinción entre el objeto y el fondo es mínima, el algoritmo puede fallar en detectar y seguir el movimiento con precisión. Las texturas uniformes o con bajo contraste no proporcionan suficientes características distintivas que el algoritmo pueda utilizar para diferenciar y rastrear el objeto. Esta falta de puntos de referencia claros puede llevar a una pérdida de precisión en el seguimiento y a errores significativos en la trayectoria del objeto.

- **Reflejos y sombras.** Los reflejos y las sombras representan otro desafío significativo para el algoritmo RLOF. Estos fenómenos pueden crear artefactos visuales que confunden al algoritmo, dificultando su capacidad para mantener un seguimiento preciso del objeto. Los reflejos pueden hacer que una superficie parezca diferente de lo que realmente es, mientras que las sombras pueden oscurecer partes del objeto, ambos factores pueden resultar en una identificación errónea de los puntos característicos o en la pérdida del objeto de interés. Estas interferencias visuales pueden requerir ajustes manuales para mantener la precisión del seguimiento.

Titularidad o Patentes

La tecnología de flujo óptico robusto local (RLOF) en OpenCV es una implementación de código abierto que forma parte de la biblioteca OpenCV, desarrollada y mantenida inicialmente por Intel y actualmente por la Fundación OpenCV. No es una tecnología propietaria ni patentada en sí misma, sino que forma parte de los componentes libres y abiertos de OpenCV.

OpenCV, incluyendo RLOF, está disponible bajo la Licencia Apache 2.0. Esto significa que los usuarios tienen un derecho no exclusivo a usar las patentes que puedan estar cubiertas en el código aportado bajo esta licencia. Además, si algún contribuyente hace valer reclamaciones de patentes contra el software cubierto por la licencia Apache 2.0, perderá su derecho a usar otras patentes en el software

- **Enlace a la Licencia Apache 2.0**: Apache License 2.0
- **Repositorio y licencia de OpenCV**: OpenCV GitHub License
-

Políticas de Privacidad para Desarrollo Local de OpenCV

Estas políticas de privacidad se aplican a los desarrollos realizados localmente utilizando la biblioteca OpenCV, específicamente en relación con la tecnología RLOF (Robust Local Optical Flow). A continuación, se describen las prácticas recomendadas para el manejo de datos y la privacidad cuando se trabaja en un **entorno local.**

1. Recopilación de Datos

En un entorno de desarrollo local, OpenCV no recopila datos de usuario automáticamente. Todas las operaciones y el procesamiento de datos se realizan localmente en su sistema, lo que garantiza que los datos no se transmiten a terceros ni se almacenan en servidores externos.

2. Uso de Datos

Los datos de imagen y video utilizados en el proyecto se procesan exclusivamente en el entorno local. Los datos pueden ser almacenados temporalmente en el sistema local durante el desarrollo y las pruebas.

3. Actualizaciones de Seguridad

Es crucial mantener el software OpenCV y los sistemas operativos actualizados con los últimos parches de seguridad para prevenir vulnerabilidades.

4. Publicación y Distribución de Datos

Si se decide compartir el código o datos de proyectos fuera del entorno local, considere anonimizar o eliminar cualquier dato sensible antes de la distribución.

5. Derechos del Usuario

El desarrollador tiene el control total sobre los datos que utiliza y genera durante el desarrollo local. Esto incluye la capacidad de modificar, eliminar o transferir dichos datos según sea necesario. Al trabajar localmente, es importante asegurar el cumplimiento con las leyes y regulaciones locales de protección de datos aplicables.

Disponibilidad y Costes de Implantación

- **Infraestructura de Hardware Necesaria**

La implementación de RLOF (Robust Local Optical Flow) en OpenCV es eficiente y puede ejecutarse en una amplia gama de hardware. Sin embargo, para obtener un rendimiento óptimo, se recomienda disponer de una infraestructura con ciertas características específicas. A continuación, se detallan las recomendaciones de hardware necesarias para maximizar el rendimiento de RLOF en OpenCV:

- **CPU**: Se recomienda un procesador moderno con múltiples núcleos y soporte para instrucciones SIMD (Single Instruction, Multiple Data). Un ejemplo adecuado sería el AMD Ryzen 5 o cualquier procesador de una gama superior. Estos procesadores ofrecen la capacidad de manejar múltiples hilos de ejecución simultáneamente, lo cual es crucial para procesar datos de manera eficiente y rápida.
- **GPU**: Para acelerar significativamente el procesamiento, es altamente recomendable contar con una tarjeta gráfica dedicada compatible con OpenCL o CUDA. Las tarjetas de la serie NVIDIA RTX, conocidas por su alto rendimiento en tareas de cómputo paralelo, son una excelente opción. También, las tarjetas de la gama profesional como las NVIDIA A100 son ideales para aplicaciones más exigentes, proporcionando una capacidad de procesamiento superior y optimizaciones específicas para cargas de trabajo intensivas.
- **Memoria RAM:** Disponer de al menos 16 GB de memoria RAM es fundamental, especialmente si se trabaja con videos de alta resolución o se desea paralelizar parte del proceso computacional. Una mayor cantidad de RAM permite manejar volúmenes de datos más grandes y mejorar la eficiencia del procesamiento, reduciendo los tiempos de espera y permitiendo una experiencia de usuario más fluida.
- **Consideraciones Adicionales:** Es esencial tener instalada la última versión de OpenCV compatible con el sistema operativo y el hardware en uso. OpenCV recibe actualizaciones frecuentes que incluyen optimizaciones y mejoras de rendimiento, por lo que mantener el software actualizado es crucial. Además, se recomienda aprovechar las optimizaciones específicas de hardware, como el uso de la GPU, para mejorar el rendimiento del RLOF. Esto incluye configurar correctamente los entornos de desarrollo y ejecutar pruebas para asegurarse de que las optimizaciones están siendo utilizadas de manera efectiva.

Fuentes de ayuda y formación

- **Documentación oficial de OpenCV.** La documentación oficial de OpenCV es una fuente completa y exhaustiva que proporciona detalles técnicos sobre la implementación y uso de la biblioteca de visión artificial. Esta documentación incluye descripciones detalladas de los algoritmos, explicaciones sobre los parámetros y configuraciones, así como numerosos ejemplos de código que demuestran cómo utilizar las diferentes funcionalidades de OpenCV. La documentación está diseñada tanto para principiantes como para usuarios avanzados, ofreciendo una guía clara y estructurada para la correcta implementación de técnicas de visión por computadora.

- **Foros y Comunidades de OpenCV.** Los foros y comunidades en línea dedicadas a OpenCV son recursos valiosos para obtener ayuda y soporte de otros desarrolladores y expertos en la materia. En estos espacios, los usuarios pueden participar en discusiones sobre problemas comunes, compartir soluciones, y aprender sobre las mejores prácticas en la implementación de algoritmos de visión por computadora. Estos foros son particularmente útiles para resolver dudas específicas, obtener recomendaciones sobre cómo optimizar el uso de OpenCV, y mantenerse actualizado con las últimas novedades y actualizaciones de la biblioteca.

- **Repositorio de código.** El repositorio oficial de OpenCV en GitHub es otra fuente crucial de información y recursos. Aquí se encuentra el código fuente del proyecto, junto con ejemplos prácticos y documentación técnica detallada. Este repositorio no solo permite a los usuarios estudiar y comprender cómo se implementan los algoritmos en OpenCV, sino que también ofrece la oportunidad de reportar problemas, sugerir mejoras y contribuir al desarrollo continuo del proyecto.

Stable Diffusion

Introducción

Stable Diffusion es una tecnología que fue lanzada en Agosto de 2022. Se desarrolló inicialmente como un modelo de generación de imágenes basado en redes neuronales de difusión, una técnica avanzada de inteligencia artificial que permite crear imágenes de alta calidad a partir de descripciones textuales.

En octubre de 2022, Stable Diffusion 1.5 fue lanzada, como una actualización del modelo original de Stable Diffusion. Esta versión introdujo mejoras significativas en la calidad de las imágenes generadas y en la eficiencia del proceso de difusión, pero manteniendo una resolución de 512x512

Stable Diffusion XL fue lanzada en julio de 2023 como una evolución significativa del modelo de generación de imágenes mediante IA. Esta versión fue diseñada para superar las limitaciones de sus predecesoras, ofreciendo una capacidad superior para generar imágenes de alta resolución(1024x1024) con un nivel de detalle y coherencia sin precedentes. Stable Diffusion XL se ha consolidado como una herramienta avanzada en la industria creativa y tecnológica, expandiendo su utilidad en campos que van desde el diseño gráfico hasta la producción de efectos visuales (VFX).

Stable Diffusion está desarrollada por Stability AI, una empresa emergente en el campo de la inteligencia artificial, en colaboración con otras entidades de investigación como EleutherAI y LAION. Stability AI está formada por un equipo de expertos en machine learning y gráficos computacionales, con una amplia experiencia en el desarrollo de modelos de inteligencia artificial de última generación. Además de trabajar en Stable Diffusion, la compañía también se dedica a proyectos de inteligencia artificial enfocados en áreas como el procesamiento del lenguaje natural y la automatización de tareas creativas. La empresa ha recibido financiación significativa de inversores privados y fondos de investigación, lo que les ha permitido seguir innovando en este campo.

Stable Diffusion permite la generación de imágenes de alta calidad a partir de descripciones textuales, utilizando un modelo de difusión que es capaz de capturar y replicar detalles complejos y estilos artísticos variados. Esta funcionalidad es especialmente relevante en la industria de VFX, donde se utiliza para generar conceptos visuales rápidos, diseñar escenas complejas y experimentar con diferentes estilos visuales antes de su implementación final en proyectos cinematográficos y de videojuegos. La capacidad de Stable Diffusion para producir imágenes realistas y estilizadas a partir de texto lo convierte en una herramienta poderosa para artistas y desarrolladores en la industria creativa.

Descripción

Stable Diffusion XL es una poderosa herramienta de generación de imágenes mediante inteligencia artificial. Esta tecnología destaca en la industria de efectos

visuales (VFX) por su capacidad de generar imágenes realistas y detalladas a partir de descripciones textuales. Esto la hace especialmente útil para artistas y estudios que buscan optimizar la creación de contenido visual de alta calidad.

La estrategia de Stability de ofrecer este modelo bajo la licencia Creative ML OpenRAIL-M license, que permite su uso comercial y no comercial con mínimas restricciones ha provocado un crecimiento exponencial del ecosistema del modelo. De esta manera el uso de Stable Diffusion XL y su ecosistema permite un gran control y flexibilidad a la hora de generar o modificar imágenes además de permitir un uso comercial de todas ellas.

A continuación, se describen las tres características más destacables:

1. Generación de Imágenes de Alta Resolución

Stable Diffusion XL permite crear imágenes de alta calidad con una resolución superior, lo que es fundamental en la producción de efectos visuales que requieren detalles precisos y realistas. Además, la comunidad de usuarios ha generado flujos de trabajo que permiten usar el modelo para escalar las imágenes generadas a 4K incrementando el nivel de detalle durante el proceso de escalado (upscaling)

2. Control Avanzado de la Generación de Imágenes

Como se indica en la introducción de esta sección, el ecosistema de Stable Diffusion XL ha crecido a una velocidad vertiginosa y proporciona múltiples herramientas para poder tener un control avanzado de las imágenes generadas por el modelo. Estas herramientas incluyen:

- Modelos reentrenados facilitan la generación de imágenes más fotorrealistas como, Realistic Vision v6.0 o Juggernaut XL
- LORAs que permiten controlar el estilo como Midjourney-Mimic
- Control Nets que permiten controlar la composición de la imagen mediante la definición de contornos, mapas de profundidad y poses como Controlnet-union-sdxl
- Image Prompting Adapters que permiten controlar el estilo y la composición a partir de una sola imagen, como IPAdapter
- Versiones del modelo entrenados para inpaint, permiten usar el modelo solo para rellenar partes de una imagen existente como Stable Diffusion XL 1.0 Inpainting

Todas estas mejoras generadas sobre la arquitectura del modelo de Stable Diffusion XL ofrecen una gran flexibilidad en la personalización de la creación y relleno de images, permitiendo a los artistas adaptar la apariencia de las imágenes generadas para cumplir con las necesidades específicas de cada proyecto.

3. Optimización de Recursos y Tiempo

Stable Diffusion XL proporciona una buen equilibrio entre calidad, tiempos de inferencia y recursos necesarios para ejecutarse. El modelo consta de 2.6 billones de parámetros tal y como se explica en el paper SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis, esto implica un consumo alrededor de 7 GB de memoria de tarjeta gráfica. Hay que tener en cuenta que si se va a utilizar herramientas generadas por el ecosistema para mejorar el control de la generación de la imagen como control nets y LORAs, el consumo de memoria de tarjeta gráfica se puede elevar al doble.

Además los tiempos de generación de imágenes usando parámetros standard y una configuración de *workstation* profesional con una NVIDIA RTX 4090 está entorno a los 2-3 segundos por imagen de 768x768 como se muestra en el estudio de Tom´s hardware Stable Diffusion Benchmarks: 45 Nvidia, AMD, and Intel GPUs Compared.

Esta combinación de calidad, tiempos de inferencia y recursos hace de Stable Diffusion XL y su ecosistema una herramienta que permite reducir significativamente el tiempo y los recursos necesarios para generar efectos visuales complejos, lo que permite a los estudios de VFX aumentar su productividad y eficiencia.

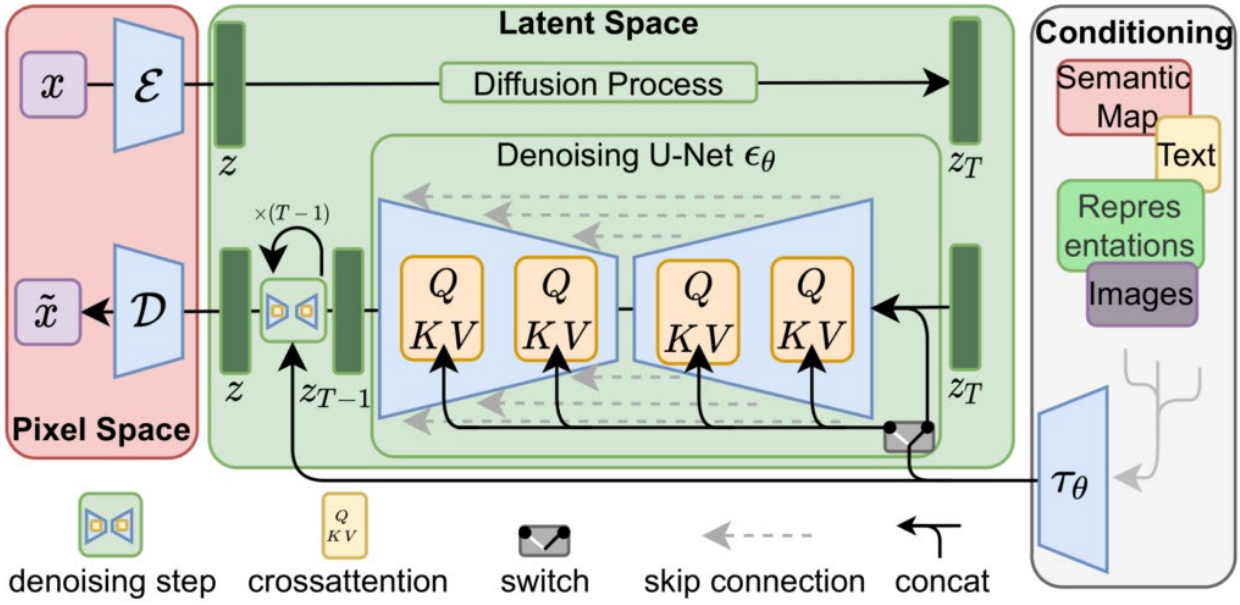


Fig 27. Esquema de la arquitectura base de Stable Diffusion

Casos de Uso

Desde el contexto de las técnicas incluidas en el radar tecnológico, el uso de Stable Diffusion XL y todo su ecosistema puede proporcionar mejoras de calidad y costes a la hora de limpiar y corregir áreas de una imagen rellenandolas con nuevos contenidos que puedan describirse de forma sencilla con texto. También puede

proporcionar beneficios a la hora de crear propuestas de fondos de planos que luego pueden ser desarrollados con más detalle por artistas de matte painting. La limitación en resolución, si bien puede ser gestionada usando técnicas de upscaling, plantea un reto a la hora de producir imágenes o contenidos a altas resoluciones como 4K que ya son casi estándar en la industria.

Casos de uso ideales

- Limpieza y preparación de planos

  Usando Stable Diffusion XL entrenado para tareas de inpainting en colaboración con control nets y IPAdapter permiten seleccionar áreas de imágenes con una máscara y rellenar este área con contenido que se integra con el resto de la imágen. Esta labor se realiza de forma rápida y el modelo puede ofrecer múltiples opciones y el artista puede elegir de entre ellas cuál se ajusta más a sus necesidades.

- Creación rápida de prototipos visuales

  Permite generar conceptos visuales preliminares para efectos visuales en minutos. Ideal para crear paisajes, escenarios o personajes que sirvan como base para proyectos VFX. También es muy útil para explorar diferentes estilos y composiciones visuales durante la fase de preproducción. El uso de herramientas como IPAdapter yo LORAs permiten explorar múltiples estilos de una manera muy ágil pero controlada.

Limitaciones

- Generación de contenidas en muy alta resolución

Como ya hemos comentado el modelo de Stable Diffusion es capaz de generar imágenes 1K con las siguientes resoluciones, tal y como se especifica en las recomendaciones que proporcionadas por Stability su documentación SDXL AWS Documentation

Fullscreen: 4:3 - 1152x896

Widescreen: 16:9 - 1344x768

Ultrawide: 21:9 - 1536x640

Mobile landscape: 3:2 - 1216x832

Square: 1:1 - 1024x1024

Mobile Portrait: 2:3 - 832x1216

Tall: 9:16 - 768x1344

El uso de otras resoluciones puede generar artefactos. Por esta razón, si el contenido que necesitamos generar es de una resolución mayor, será necesario hacerlo en dos pasos:

1. Generación a baja resolución
2. Upscaling a la resolución objetivo

Desafortunadamente, el segundo paso suele requerir un mayor uso de memoria de tarjeta gráfica y suele requerir minutos en vez de segundos para completarse. Además, complica el flujo de trabajo

- Generación de imágenes con estéticas no fotorrealistas

  El ecosistema actual de Stable Diffusion XL proporciona múltiples herramientas para generar imágenes fotorealistas o de estética de cómic japonés. Sin embargo, puede resultar más complicado generar contenido no fotorrealista con estéticas muy estilizadas. Para ello sería necesario entrenar un LORA propio. Esta labor fine tuning de modelos para poder ajustarlos a las necesidades de un proyecto no es una labor que cada vez es más asequible pero requiere de unos conocimientos y que pueden estar fuera del ámbito de la industria de VFX y por tanto involucran una inversión extra en talento y formación.

Titularidad o Patentes

Stable Diffusion XL es una tecnología de código abierto desarrollada por Stability AI. Aunque el modelo es accesible al público, su uso está regulado por términos y condiciones específicos.

Licencia de uso

Stable Diffusion XL se distribuye bajo una licencia de código abierto que permite a los usuarios descargar, modificar y utilizar el modelo para propósitos personales y comerciales, con ciertas restricciones. No se permite el uso del modelo para generar imágenes que promuevan odio, violencia o cualquier actividad ilegal. Además, los usuarios deben atribuir adecuadamente la tecnología cuando la utilicen en sus proyectos.

Licencia de Uso de Stable Diffusion

Política de privacidad

Política de Privacidad de Stability AI

Disponibilidad y Costes de Implantación

Stable Diffusion XL está disponible bajo un modelo de código abierto, lo que significa que es gratuito para su uso general.

Infraestructura Hardware

Para ejecutar Stable Diffusion XL eficientemente, se requiere una infraestructura de hardware considerable:

- **Tarjetas Gráficas (GPU):** Se recomienda una GPU moderna con al menos 12 GB de VRAM, como las series NVIDIA RTX 30 o superiores.
- **Memoria RAM:** Mínimo 16 GB, aunque 32 GB o más es ideal para manejar grandes volúmenes de datos.
- **CPU:** Procesadores multi-núcleo con alto rendimiento, como Intel i7/i9 o AMD Ryzen 7/9.
- **Almacenamiento SSD:** Al menos 1TB GB de espacio libre en disco para manejar el modelo y los archivos generados.

Fuentes de Ayuda y Formación

Para facilitar la adopción de Stable Diffusion XL, existen múltiples recursos de ayuda y formación:

- **Documentación Oficial:** Stability AI ofrece una documentación detallada disponible Huggingface.
- **Comunidades en línea:** Plataformas como GitHub, Reddit, y Discord tienen comunidades activas donde los usuarios pueden intercambiar conocimientos, resolver dudas y compartir recursos.
- **Tutoriales y Cursos en línea:** Hay numerosos tutoriales gratuitos y cursos en plataformas como YouTube, Coursera y Udemy que enseñan cómo instalar, ejecutar, y personalizar Stable Diffusion XL.
- **Soporte Técnico:** Stability AI ofrece soporte técnico a través de su portal oficial, así como opciones de soporte avanzado para clientes empresariales si se usa el modelo a través de su API.

Plataformas

Tras un análisis de las plataformas actuales disponibles, se incluyen en este documento las que se consideran más adecuadas para el tipo de tecnologías disruptivas que se analizan en este radar tecnológico.

NVIDIA RTX 4090

Introducción

La **NVIDIA RTX 4090** es una tarjeta gráfica de alta gama perteneciente a la serie GeForce RTX 40, desarrollada por **NVIDIA Corporation**, una empresa líder en el diseño de unidades de procesamiento gráfico (GPU). Esta tarjeta fue lanzada al mercado en octubre de 2022 y está dirigida principalmente a usuarios que requieren un rendimiento alto, como creadores de contenido, y profesionales en áreas como la inteligencia artificial. La RTX 4090 utiliza la arquitectura **Ada Lovelace**, que representa la última generación en tecnologías de gráficos y cómputo acelerado.

La serie GeForce RTX se introdujo por primera vez en 2018, inaugurando una nueva era en gráficos de computadora al ser las primeras GPUs en incorporar trazado de rayos en tiempo real (ray tracing) y aceleración por inteligencia artificial (IA) mediante núcleos Tensor. La NVIDIA RTX 4090, lanzada en octubre de 2022, es el modelo más potente de la serie 40, y continúa la evolución de estas tecnologías, mejorando drásticamente el rendimiento y la eficiencia en comparación con generaciones anteriores.

NVIDIA Corporation, fundada en 1993, es la empresa detrás del desarrollo de la RTX 4090. NVIDIA es una empresa pionera en el diseño de unidades de procesamiento gráfico (GPU) y ha sido fundamental en la evolución de la computación gráfica y el desarrollo de tecnologías para juegos, inteligencia artificial y computación de alto rendimiento (HPC). La empresa cuenta con un equipo de ingenieros y científicos de alto nivel que continuamente innovan en el campo de los gráficos por computadora. NVIDIA, con sede en Santa Clara, California, es también conocida por su sólido respaldo financiero, lo que le permite realizar grandes inversiones en investigación y desarrollo.

La NVIDIA RTX 4090 es una tarjeta gráfica que destaca por su capacidad en inteligencia artificial (IA) y renderizado avanzado. Gracias a su arquitectura Ada Lovelace y a la integración de núcleos Tensor de cuarta generación, la RTX 4090 es capaz de acelerar significativamente las operaciones de IA, lo que es crucial para la adopción e implementación de todas las técnicas incluídas en el radar como el entrenamiento y la inferencia de modelos de aprendizaje profundo. Además, incorpora núcleos RT (Ray Tracing) mejorados que permiten un trazado de rayos en tiempo real extremadamente realista, lo cual es esencial para el renderizado de gráficos en 3D de alta fidelidad.

Descripción

La RTX 4090 se destaca por sus avanzadas capacidades en inteligencia artificial, renderizado y relación calidad-precio. Con 512 núcleos Tensor de cuarta generación, ofrece una aceleración excepcional de las operaciones de multiplicación de matrices ofreciendo un alto rendimiento en las tareas de inferencia de modelos como Segment Anything, Stable Diffusion y YOLO. Su precio competitivo frente a la NVIDIA RTX A6000 la convierte en una opción atractiva para profesionales que buscan alto rendimiento a un costo más accesible.

A continuación, se describen las tres características más destacables:

- Aceleración de Inteligencia Artificial (IA)

  La RTX 4090 cuenta con 512 **núcleos Tensor de cuarta generación**, que están optimizados para acelerar tareas relacionadas con inteligencia artificial, como el entrenamiento y la inferencia de modelos de aprendizaje profundo.

  Los núcleos Tensor permiten acelerar la multiplicación de matrices 2x. Estos núcleos son tan rápidos que el cuello de botella pasa a ser la velocidad de transferencia de la memoria VRAM a las cachés L1 y L2 de la tarjeta. La RTX 4090 goza también de una un ancho de banda de memoria de 1008 GB/s además de una L2 de 72MB.

  Esta capacidad es crucial para casos de uso de IA que requieren un alto nivel de procesamiento paralelo, permitiendo la ejecución rápida y eficiente de modelos de IA incluídos en este radar tecnológico.

- Precio Competitivo Comparado con la NVIDIA RTX A6000

  La RTX 4090 ofrece un **precio competitivo** en comparación con la NVIDIA RTX A6000, especialmente teniendo en cuenta su rendimiento similar o superior en muchas aplicaciones de IA y renderizado. Aunque la A6000 está diseñada específicamente para estaciones de trabajo y entornos profesionales de alta disponibilidad y rendimiento, la RTX 4090 proporciona una alternativa más accesible sin sacrificar capacidades clave, lo que la convierte en una opción atractiva.

- Trazado de Rayos en Tiempo Real (*Ray Tracing*)

Equipado con **núcleos RT de tercera generación**, la RTX 4090 ofrece capacidades avanzadas de trazado de rayos en tiempo real, lo que resulta en un renderizado extremadamente realista de gráficos 3D. Esto es especialmente relevante en

metodologías de renderizado que buscan simular con precisión el comportamiento de la luz y las sombras, mejorando la calidad visual en aplicaciones como VFX.

Fig 28. NVIDIA RTX 4090

Casos de Uso

La RTX 4090 es una tarjeta gráfica de alto rendimiento que destaca por sus 512 núcleos Tensor de cuarta generación, optimizados para acelerar tareas de inteligencia artificial como el entrenamiento y la inferencia de modelos de aprendizaje profundo. Ofrece un rendimiento excepcional a un precio competitivo. Su única potencial limitación es la VRAM de 24 GB y su consumo de energía y sistema de refrigeración que no ha sido diseñado para entornos de alta disponibilidad y utilización.

- **Casos de Uso Ideales**

1. Entrenamiento e Inferencia de Modelos de Inteligencia Artificial (IA)

Ideal para inferenciar modelos de aprendizaje profundo con un número de parámetros que no use más de 24GB. Gracias a sus núcleos Tensor de cuarta generación. Es excelente para aplicaciones de IA que requieren procesamiento masivo, como redes neuronales convolucionales (CNNs), procesamiento de lenguaje natural (NLP), y visión por computadora. En el informe elaborado por Tom´s hardware en diciembre de 2023, la RTX 4090 es la tarjeta más rápida generando imágenes con Stable Diffussion.

2. Renderizado de Gráficos 3D y Animación

Perfecta para estudios de VFX y animación que necesitan acelerar el renderizado de escenas complejas con trazado de rayos en tiempo real. En el informe elaborado por Tom´s hardware en Agosto de 2024, sobre las mejores tarjetas gráficas para gaming,

que correlaciona muy bien con el rendimiento de renders offline como Arnold, VRAY y Octane, la RTX 4090 es la tarjeta con mejor rendimiento.

- **Limitaciones**

1. Memoria VRAM Inferior a la NVIDIA RTX A6000

La RTX 4090 tiene **24 GB de VRAM**, lo cual es menor que los **48 GB** de la NVIDIA RTX A6000. Esta diferencia en capacidad de memoria puede limitar el rendimiento de pipelines de ComfyUI o Diffusers que usen varios modelos de IA con un gran número de parámetros que en total requieran de más de 24 GB de VRAM..

2. Aplicaciones en Entornos de Cómputo de Alto Rendimiento (HPC)

No es la mejor opción para entornos de cómputo de alto rendimiento (HPC) que requieren el uso continuo de la tarjeta gráfica sin incrementar el riesgo de rotura del hardware, una memoria gráfica muy grande, un consumo y refrigeración sostenible..

Titularidad o Patentes

La tecnología detrás de la NVIDIA RTX 4090es propietaria y está protegida por diversas patentes y derechos de propiedad intelectual de NVIDIA Corporation. Esto incluye tanto el diseño de hardware como el software asociado, como los controladores y tecnologías de aceleración específicas (por ejemplo, CUDA, Tensor Cores, RT Cores). Las patentes y derechos de propiedad intelectual aseguran que NVIDIA mantenga el control exclusivo sobre el desarrollo, fabricación y distribución de esta tecnología. Los usuarios deben cumplir con los términos de uso establecidos por NVIDIA para evitar infracciones legales.

Licencia de Uso

El uso de la **NVIDIA RTX 4090** está sujeto a los términos y condiciones establecidos por NVIDIA, que regulan el uso del hardware y el software asociado. Esto incluye la prohibición de la ingeniería inversa, la distribución no autorizada, y el uso en aplicaciones no permitidas por la licencia.

- Link a la licencia de venta
- Eula CUDA

Política de Privacidad

NVIDIA recopila y utiliza datos de los usuarios de acuerdo con su política de privacidad, que cubre aspectos como la recopilación de datos personales, el uso de esos datos, y la protección de la privacidad del usuario. Esto incluye información recogida durante el registro de productos, el uso de software y servicios en línea proporcionados por NVIDIA.

- Link a la Política de Privacidad

Disponibilidad y Costes de Implantación

Los precios de la **NVIDIA RTX 4090** varían según la región y el proveedor, pero en general, esta tarjeta gráfica se encuentra en el rango de **1,800 a 2,500 EUR**. Este coste incluye la licencia de uso estándar para la tarjeta gráfica, que generalmente cubre su utilización en entornos de trabajo individuales o empresariales.

- **Licencia de Usuario Final (EULA):** Incluida en el precio de compra del hardware. No tiene costes adicionales para su uso básico.
- **Licencia para Centros de Datos:** Puede haber costos adicionales si se requiere soporte especializado o servicios adicionales de virtualización.
- **Licencia para Desarrolladores:** Aunque la tarjeta viene con la EULA estándar, los desarrolladores que deseen distribuir aplicaciones desarrolladas con el uso de la RTX 4090 podrían necesitar adquirir licencias adicionales para software de desarrollo como SDKs o APIs específicos.

Tarjeta Gráfica	Arquitectura	CUDA Núcleos	Tensor Cores	VRAM (GB) GDDR6	Ancho de banda memoria (GB/s)	Precio Orientativo (EUR)
NVIDIA RTX 4090	Ada	16.384	512	24	1008	€1.779,00

Infraestructura Hardware Necesaria:

La **NVIDIA RTX 4090** requiere una infraestructura de hardware avanzada para funcionar de manera óptima:

- **Unidad Central de Procesamiento (CPU):** Procesadores de alta gama como Intel Xeon o AMD Ryzen Threadripper son recomendados para manejar el procesamiento paralelo con la GPU.
- **Fuente de Alimentación:** Se requiere una fuente de alimentación de al menos 850W, con conectores PCIe de 16 pines dedicados.
- **Espacio de Enfriamiento:** La RTX 4090 puede generar una cantidad significativa de calor, por lo que es esencial contar con un sistema de enfriamiento adecuado, ya sea por aire o líquido.
- **Chasis de Ordenador:** Se necesita un chasis que soporte tarjetas gráficas de gran tamaño, asegurando espacio suficiente para la correcta instalación y ventilación.

Fuentes de Ayuda y Formación:

Para asegurar la adopción exitosa de la **NVIDIA RTX 4090**, NVIDIA y otras fuentes ofrecen diversos recursos de ayuda y formación:

- **Documentación Técnica y Manuales:** NVIDIA proporciona una amplia gama de manuales y guías técnicas que cubren desde la instalación hasta la optimización del rendimiento.
- **NVIDIA Developer Program:** Un programa diseñado para desarrolladores que incluye acceso a foros, webinars, SDKs, y ejemplos de código para maximizar el uso de la tecnología.
- **Cursos en Línea y Webinars:** NVIDIA ofrece cursos en línea y webinars a través de su plataforma NVIDIA Deep Learning Institute (DLI), enfocados en la aplicación de la RTX 4090 en áreas como IA, renderización, y computación de alto rendimiento.
- **Soporte Técnico Especializado:** Disponible a través de contratos de soporte, que ofrecen asistencia directa de ingenieros de NVIDIA para la resolución de problemas específicos.
- **Comunidad de Usuarios:** Foros y comunidades en línea donde los usuarios de la RTX 4090 pueden compartir experiencias, consejos y soluciones a problemas comunes.

NVIDIA RTX A6000

Introducción

La NVIDIA RTX A6000 es una tarjeta gráfica de alta gama diseñada para profesionales que requieren un rendimiento gráfico y de computación muy alto, como desarrolladores de inteligencia artificial, investigadores en ciencias computacionales y creadores de contenido. Esta tarjeta forma parte de la línea RTX, conocida por sus capacidades avanzadas en trazado de rayos en tiempo real y computación acelerada por GPU.

La NVIDIA RTX A6000 fue lanzada en octubre de 2020, como parte de la serie RTX Ampere de NVIDIA. Esta serie representa la segunda generación de tarjetas RTX, sucediendo a la arquitectura Turing. La arquitectura Ampere, en la que se basa la A6000, ofrece mejoras significativas en rendimiento, eficiencia energética y capacidad de memoria, respondiendo a la creciente demanda de gráficos de alta resolución y procesamiento masivo de datos. En 2022 se lanzó una nueva versión RTX A6000 usando la nueva arquitectura Ada

NVIDIA Corporation, fundada en 1993, es la empresa detrás del desarrollo de la RTX A6000. NVIDIA es una empresa pionera en el diseño de unidades de procesamiento

gráfico (GPU) y ha sido fundamental en la evolución de la computación gráfica y el desarrollo de tecnologías para juegos, inteligencia artificial y computación de alto rendimiento (HPC). La empresa cuenta con un equipo de ingenieros y científicos de alto nivel que continuamente innovan en el campo de los gráficos por computadora. NVIDIA, con sede en Santa Clara, California, es también conocida por su sólido respaldo financiero, lo que le permite realizar grandes inversiones en investigación y desarrollo.

La NVIDIA RTX A6000 está diseñada para manejar cargas de trabajo extremadamente intensivas en gráficos y cómputo. Con 48 GB de memoria GDDR6 y una arquitectura Ampere mejorada que cuenta con 10752 núcleos CUDA y 336 núcleos Tensor, la A6000 es ideal para tareas de entrenamiento e inferencia de modelos de IA, y renderizado en 3D en tiempo real. Su alta capacidad de memoria, le permite cargar modelos complejos de IA para inferencia, y permite también soportar las altas necesidades de memoria requeridas para entrenar modelos. Su configuración del CUDA y Tensor cores también la dota de una capacidad de cómputo muy competitivas.

Descripción

La funcionalidad de la NVIDIA RTX A6000 destaca por su amplia capacidad de memoria de 48 GB GDDR6, que le permite trabajar con modelos de IA complejos haciendo posible el uso eficiente de múltiples modelos de en workflows de ComfyUI o Diffusers pipelines. También tiene un rendimiento competitivo a la hora de inferenciar y entrenar modelos de IA gracias a su computación acelerada mediante Tensor Cores, optimizada para tareas de inteligencia artificial y aprendizaje automático. Por último, también destaca por su capacidad de trazado de rayos en tiempo real, que proporciona visuales extremadamente realistas en aplicaciones de diseño y simulación.

A continuación, se describen las tres características más destacables:

- Amplia Capacidad de Memoria (48 GB GDDR6 VRAM)

  La NVIDIA RTX A6000 se destaca por su enorme capacidad de 48 GB de memoria GDDR6, que es fundamental para aplicaciones avanzadas de inteligencia artificial. Esta gran cantidad de memoria permite cargar y gestionar múltiples modelos de IA simultáneamente, lo que es crucial para técnicas complejas como la limpieza de planos asistida por IA generativa. Estas aplicaciones requieren que los modelos estén completamente cargados en la memoria de la tarjeta para asegurar un rendimiento competitivo y una rapidez en la regeneración de imagenes y máscaras. Aunque también es capaz de manejar tareas como la renderización de escenas en 3D y la edición de vídeo en 8K, es en el ámbito de la inteligencia artificial donde la capacidad de memoria de la RTX A6000 realmente se distingue, posicionándose como una herramienta esencial para profesionales que buscan ejecutar procesos de IA avanzados sin comprometer la velocidad ni la eficiencia.

- Computación Acelerada con Tensor Cores (AI-Enhanced Computing):

Equipado con 336 núcleos Tensor de tercera generación, la RTX A6000 está optimizada para tareas de aprendizaje automático y aplicaciones de inteligencia artificial. Estos núcleos permiten realizar cálculos de precisión mixta, lo que acelera significativamente el entrenamiento y la inferencia de modelos de IA.

Los núcleos Tensor permiten acelerar la multiplicación de matrices 2x. Los núcleos Tensor son tan rápidos que el cuello de botella pasa a ser la velocidad de transferencia de la memoria VRAM a las cachés L1 y L2 de la tarjeta. La RTX A6000 goza también de una un ancho de banda de memoria de 768GB/s además de una L2 de 6MB para la versión de arquitectura Ampere y 72MB para la versión A6000 Ada.

Esta característica es especialmente valiosa para su uso en la inferencia de modelos como Stable Diffusion, SAM y YOLO. Aunque su rendimiento es competitivo, la RTX 4090 y RTX A6000 Ada la superan. Si el rendimiento de la RTX A6000 no llegara al nivel necesario por los requisitos, se puede optar por la versión Ada que incluye 568 núcleos Tensor.

- Capacidad de Trazado de Rayos en Tiempo Real (Real-Time Ray Tracing)

La NVIDIA RTX A6000 cuenta con núcleos RT de segunda generación, que permiten realizar trazado de rayos en tiempo real con una gran precisión. Esta funcionalidad es crucial para profesionales en el ámbito del diseño gráfico, animación y simulación, ya que ofrece una representación extremadamente realista de las interacciones entre la luz y los objetos en escenas 3D. Esto mejora la calidad visual de los renders permite maximizar la utilización de esta tarjeta que cuando no se use para acelerar la inferencia o entrenamiento de modelos de IA puede usarse para acelerar el cálculo de renders.

Estas tres funcionalidades hacen que la NVIDIA RTX A6000 sea una herramienta indispensable para profesionales que requieren rendimiento gráfico de alta gama y capacidad de cómputo avanzada en sus proyectos de IA.

Fig 29. NVIDIA RTX A6000

Casos de Uso

La **NVIDIA RTX A6000** está diseñada específicamente para abordar las necesidades más exigentes en los campos de la inteligencia artificial y la renderización gráfica avanzada. Esta tarjeta gráfica es ideal para profesionales que trabajan con modelos de IA complejos, renderización 3D en tiempo real y otras tareas que requieren una combinación de potencia computacional y capacidad de memoria excepcionales. No obstante, su elevado costo y características especializadas pueden no ser necesarias para aplicaciones menos intensivas.

- **Casos de uso ideales**

1. **Entrenamiento y Inferencia de Modelos de Inteligencia Artificial:**
    - **Entrenamiento de redes neuronales profundas:** La capacidad de manejar grandes cantidades de datos y entrenar modelos complejos con múltiples capas es clave en la creación de aplicaciones de IA avanzadas.
    - **IA Generativa y Limpieza de Planos:** Utilización en tareas de generación de contenido y limpieza de datos asistidas por IA, donde la memoria extendida permite cargar y procesar múltiples modelos simultáneamente sin pérdida de rendimiento.
    - **Procesamiento de imágenes y visión por computadora:** Despliegue de modelos para la clasificación de imágenes, detección de objetos y análisis de video en tiempo real.
2. **Renderización 3D en Tiempo Real y Animación Avanzada:**
    - **Renderización de escenas complejas con trazado de rayos:** Creación de gráficos hiperrealistas en tiempo real, crucial para la industria del cine, videojuegos y visualización arquitectónica.
    - **Producción de contenido en alta resolución:** Edición y renderizado de videos en 4K y 8K, donde se requiere un procesamiento gráfico intensivo y una gran cantidad de memoria para manejar los detalles más finos.

- **Limitaciones**

1. **Escenarios de IA que no requieren alto rendimiento o memoria extendida:**
    - **Modelos pequeños o de bajo consumo de recursos:** En proyectos de IA donde los modelos no son grandes o complejos, la capacidad extra de la RTX A6000 podría no ofrecer una ventaja significativa sobre opciones más económicas.
2. **Renderización de Baja Complejidad:**
    - **Proyectos de renderización simple o de baja resolución:** Para tareas que no requieren trazado de rayos en tiempo real ni resoluciones ultra altas, las capacidades de la A6000 podrían ser excesivas y subutilizadas.

- **Aplicaciones donde el costo-beneficio no es favorable:** En proyectos donde el presupuesto es una limitación y no se requiere el máximo rendimiento, otras soluciones más asequibles pueden ser más adecuadas.

3. **Aplicaciones fuera del ámbito gráfico o IA intensiva:**
   - **Tareas de computación general:** Usar la RTX A6000 para tareas que no requieren gráficos avanzados o procesamiento de IA sería un desperdicio de su potencial y recursos.

Titularidad o Patentes

La tecnología detrás de la NVIDIA RTX A6000 es propietaria y está protegida por diversas patentes y derechos de propiedad intelectual de NVIDIA Corporation. Esto incluye tanto el diseño de hardware como el software asociado, como los controladores y tecnologías de aceleración específicas (por ejemplo, CUDA, Tensor Cores, RT Cores). Las patentes y derechos de propiedad intelectual aseguran que NVIDIA mantenga el control exclusivo sobre el desarrollo, fabricación y distribución de esta tecnología. Los usuarios deben cumplir con los términos de uso establecidos por NVIDIA para evitar infracciones legales.

Licencia de Uso

El uso de la **NVIDIA RTX A6000** está sujeto a los términos y condiciones establecidos por NVIDIA, que regulan el uso del hardware y el software asociado. Esto incluye la prohibición de la ingeniería inversa, la distribución no autorizada, y el uso en aplicaciones no permitidas por la licencia.

- **Link a la licencia de venta**

Política de Privacidad

NVIDIA recopila y utiliza datos de los usuarios de acuerdo con su política de privacidad, que cubre aspectos como la recopilación de datos personales, el uso de esos datos, y la protección de la privacidad del usuario. Esto incluye información recogida durante el registro de productos, el uso de software y servicios en línea proporcionados por NVIDIA.

- **Link a la Política de Privacidad**

Disponibilidad y Costes de Implantación

Precios Disponibles por Tipo de Licencia:

Los precios de la **NVIDIA RTX A6000** varían según la región y el proveedor, pero en general, esta tarjeta gráfica se encuentra en el rango de **4,500 a 7,500 EUR**. Este coste

incluye la licencia de uso estándar para la tarjeta gráfica, que generalmente cubre su utilización en entornos de trabajo individuales o empresariales.

- **Licencia de Usuario Final (EULA):** Incluida en el precio de compra del hardware. No tiene costes adicionales para su uso básico.
- **Licencia para Centros de Datos:** Puede haber costos adicionales si se requiere soporte especializado o servicios adicionales de virtualización.
- **Licencia para Desarrolladores:** Aunque la tarjeta viene con la EULA estándar, los desarrolladores que deseen distribuir aplicaciones desarrolladas con el uso de la RTX A6000 podrían necesitar adquirir licencias adicionales para software de desarrollo como SDKs o APIs específicos.

Tarjeta Gráfica	Arquitectura	CUDA Núcleos	Tensor Cores	VRAM (GB) GDDR6	Precio Orientativo (EUR)
NVIDIA RTX A6000	Ampere	10.752	336	48	€4.479,25
NVIDIA RTX 6000 Ada	Ada	18.176	568	48	€7.257,90

Infraestructura Hardware Necesaria:

La **NVIDIA RTX A6000** requiere una infraestructura de hardware avanzada para funcionar de manera óptima:

- **Unidad Central de Procesamiento (CPU):** Procesadores de alta gama como Intel Xeon o AMD Ryzen Threadripper son recomendados para manejar el procesamiento paralelo con la GPU.
- **Fuente de Alimentación:** Se requiere una fuente de alimentación de al menos 700W, con conectores PCIe de 16 pines dedicados.
- **Espacio de Enfriamiento:** La RTX A6000 puede generar una cantidad significativa de calor, por lo que es esencial contar con un sistema de enfriamiento adecuado, ya sea por aire o líquido.
- **Memoria RAM:** Se recomienda al menos 64 GB de RAM para soportar tareas complejas de IA y renderización, aunque para casos más avanzados, 128 GB o más pueden ser necesarios.
- **Chasis de Ordenador:** Se necesita un chasis que soporte tarjetas gráficas de gran tamaño, asegurando espacio suficiente para la correcta instalación y ventilación.

Fuentes de Ayuda y Formación:

Para asegurar la adopción exitosa de la **NVIDIA RTX A6000**, NVIDIA y otras fuentes ofrecen diversos recursos de ayuda y formación:

- **Documentación Técnica y Manuales:** NVIDIA proporciona una amplia gama de manuales y guías técnicas que cubren desde la instalación hasta la optimización del rendimiento de la RTX A6000.
- **NVIDIA Developer Program:** Un programa diseñado para desarrolladores que incluye acceso a foros, webinars, SDKs, y ejemplos de código para maximizar el uso de la tecnología.
- **Cursos en Línea y Webinars:** NVIDIA ofrece cursos en línea y webinars a través de su plataforma NVIDIA Deep Learning Institute (DLI), enfocados en la aplicación de la RTX A6000 en áreas como IA, renderización, y computación de alto rendimiento.
- **Soporte Técnico Especializado:** Disponible a través de contratos de soporte, que ofrecen asistencia directa de ingenieros de NVIDIA para la resolución de problemas específicos.
- **Comunidad de Usuarios:** Foros y comunidades en línea donde los usuarios de la RTX A6000 pueden compartir experiencias, consejos y soluciones a problemas comunes.

AMD RX 7900 XTX

Introducción

La AMD RX 7900 XTX fue lanzada en diciembre de 2022, como parte de la serie RX 7000 de AMD. Esta serie es la primera en utilizar la arquitectura RDNA 3, diseñada para ofrecer un rendimiento superior en tareas gráficas complejas, incluyendo aquellas que utilizan inteligencia artificial (IA). La RX 7900 XTX es la tarjeta gráfica de AMD con mayores prestaciones del mercado y con el objetivo de competir con la NVIDIA RTX 4090.

AMD (Advanced Micro Devices) es la empresa detrás de la RX 7900 XTX. Fundada en 1969, AMD es una de las principales compañías en el desarrollo de procesadores y tarjetas gráficas. En los últimos años, AMD ha centrado sus esfuerzos en competir con NVIDIA en el mercado de GPUs, logrando avances significativos con sus arquitecturas RDNA y RDNA 2, que preceden a la RDNA 3.

Además de GPUs, AMD desarrolla CPUs, APUs (combinación de CPU y GPU), y soluciones para centros de datos. La financiación de AMD proviene de sus ingresos por ventas, inversiones en I+D, y su cotización en la bolsa NASDAQ. AMD ha crecido significativamente gracias a sus éxitos recientes en los mercados de computación de alto rendimiento y gaming.

La AMD RX 7900 XTX está diseñada para ofrecer un rendimiento excepcional en tareas gráficas intensivas, incluyendo aquellas que requieren inteligencia artificial. En el contexto de la industria de VFX, esta tarjeta gráfica permite acelerar procesos como el rendering en tiempo real, simulaciones de partículas y fluidos, y la creación de animaciones generadas por IA. La RX 7900 XTX cuenta con una gran cantidad de núcleos de procesamiento y memoria rápida, lo que la hace ideal para manejar grandes volúmenes de datos y operaciones de IA.

Descripción

La AMD RX 7900 XTX se destaca por su capacidad para manejar tareas gráficas intensivas, especialmente aquellas que implican el uso de inteligencia artificial (IA) y rendering en tiempo real, elementos clave en la industria de efectos visuales (VFX).

A continuación, se describen las tres características más destacables:

- Aceleración de Inferencia y Entrenamiento de modelos de IA

  La nueva arquitectura RDNA 3 de la RX 7900 XTX permite acelerar la multiplicación de matrices gracias a sus 96 núcleos que soportan WMMA. Esta funcionalidad necesita el uso del stack the librerías y drivers ROCm de AMD para poder ser utilizada. Pytorch, una de las librerías de IA más utilizadas, proporciona su

última version para Linux con ROCm permitiendo así acelerar su ejecución usando esta tarjeta gráfica.

- Capacidad y velocidad de VRAM alta

La RX 7900 XTX tiene 24GB de memoria GDDR6 con un ancho de banda de 960 GB/s. Esto permite cargar varios modelos de IA distintos a la vez permitiendo asi que su ejecución pueda aprovecharse de la aceleración de inferencia y entrenamiento que proporciona la tarjeta. También proporciona un ancho de banda de acceso a esta memoria de 960 GB/s que permite que la memoria no se convierta en un cuello de botella a la hora de maximizar el uso de los núcleos de cálculo de matrices acelerado

- Aceleración de Renderizado en Tiempo Real

La nueva arquitectura RDNA 3 de la RX 7900 XTX mejora el rendimiento del trazado de rayos permitiendo una aceleración del renderizado fotorealista, crucial para iteraciones rápidas en VFX y animación. Cuenta con 6144 Shading units y

Fig 30. AMD RX 7900 XTX

Casos de Uso

La RX 7900 XTX es una de las mejores opciones de AMD en el mercado para abordar las necesidades más exigentes en los campos de la inteligencia artificial y la renderización gráfica avanzada. Sin embargo, debido a las limitaciones en la capacidad

de integración de núcleos de AMD, no es capaz de competir con el rendimiento de NVIDIA RTX 4090 y A6000. Aunque su nueva arquitectura RDNA3 permite una gran aceleración de la multiplicación de matrices, al solo contar con 96 núcleos que soportan esta funcionalidad frente a los 512 núcleos Tensor de la NVIDIA RTX 4090, su rendimiento final no es competitivo. Además, todas las librerías relevantes de IA soportan la aceleración usando núcleos Tensor, mientras que solo algunas librerías en ciertos sistemas operativos son capaces de usar la aceleración de multiplicación de matrices de la RX7900 XTX

Casos de uso ideales

1. Inferencia de modelos con pocos requisitos de computo
2. Renderizado acelerado con bajo presupuesto

Limitaciones

1. Rendimiento por debajo de las tarjetas NVIDIA de su gama:

Como se explica en el siguiente artículo, *Recomputing gpu performance*, el rendimiento de la NVIDIA RTX 4090 es de más del doble que la RX 7900 XTX. Y esta limitación viene dada por el hardware no por el software. Aunque su precio es menor que su competidora, el uso de IA en producción de VFX necesita la máxima aceleración para poder ser competitivo con las metodologías tradicionales.

2. Bajo soporte de librerías de IA

Las librerías que permiten el uso de la aceleración de las tarjetas NVIDIA, CUDA, es ampliamente soportada por todas las librerías de IA más relevantes (TensorFlow, Pytorch, etc..) tanto en Linux como en Windows. Las librerías de AMD que permiten la aceleración hardware en tarjetas gráficas AMD, ROCm, solo es soportada por Pytorch en Linux, aunque AMD está haciendo un gran esfuerzo por cambiar esta tendencia.

3. Menor optimización y mayor tasa de bugs

Debido a la baja adopción de las librerías ROCm, los usuarios de Pytorch para AMD suelen sufrir un mayor número de bugs y el nivel de optimización es menor frente a la versión de Pytorch que soporta CUDCA

Titularidad o Patentes

La tecnología detrás de la AMD RX 7900 XTX es propiedad de **Advanced Micro Devices, Inc. (AMD)**. AMD posee patentes y derechos de propiedad intelectual sobre la arquitectura RDNA 3 y otras tecnologías integradas en la RX 7900 XTX. Esto incluye patentes relacionadas con la optimización de procesamiento gráfico, técnicas avanzadas de renderizado, y eficiencia energética.

Uso Legal y Patentes

La tecnología de AMD está protegida por diversas patentes registradas a nivel internacional. El uso y distribución de sus productos están sujetos a licencias específicas que limitan cómo pueden ser utilizados, distribuidos y modificados. Los detalles legales específicos sobre el uso de las tecnologías patentadas por AMD se pueden encontrar en los siguientes enlaces:

Licencias de Productos y Condiciones de Uso: AMD Licencias de uso

Políticas de Privacidad

AMD también cuenta con políticas de privacidad que describen cómo maneja los datos personales de sus clientes y usuarios. Estas políticas cubren la recopilación, uso, y protección de datos personales, y se pueden consultar en el siguiente enlace:

Política de Privacidad de AMD: AMD Política de Privacidad

Estas políticas aseguran que AMD cumple con las normativas internacionales de privacidad, como el Reglamento General de Protección de Datos (GDPR) en Europa, y proporcionan información detallada sobre los derechos de los usuarios en relación con sus datos.

Disponibilidad y Costes de Implantación

Licencias y Precios

La AMD RX 7900 XTX no requiere licencias de software para su funcionamiento, ya que es un componente de hardware. Sin embargo, los precios de la tarjeta gráfica varían según el fabricante y las configuraciones personalizadas ofrecidas por diferentes OEMs (Original Equipment Manufacturers) como ASUS, MSI, y Gigabyte. A partir de su lanzamiento, el precio de referencia de la RX 7900 XTX rondaba los $999 USD, aunque el costo puede variar dependiendo de la región y las características adicionales que incluyan los fabricantes.

Tarjeta Gráfica	Arquitectura	Compute Nucleos	VRAM (GB) GDDR6	Ancho de banda de memoria (GB/s)	Precio Recomendado (EUR)
AMD RX 7900 XTX	RDNA 3	96	24	960	€1,199.00

Infraestructura Hardware Necesaria

Para aprovechar al máximo la AMD RX 7900 XTX, es necesario contar con una infraestructura de hardware adecuada:

- **Procesador (CPU):** Un procesador de alto rendimiento, preferiblemente de la serie AMD Ryzen 5000 o superior, o un Intel Core i7/i9 de última generación.
- **Memoria (RAM):** Al menos 16 GB de RAM, aunque se recomienda 32 GB o más para tareas intensivas en VFX.
- **Fuente de Alimentación:** Una fuente de alimentación de al menos 850W con certificación 80 Plus Gold o superior, para garantizar un suministro de energía estable.
- **Sistema de Refrigeración:** Sistemas de enfriamiento avanzados, como refrigeración líquida, para mantener las temperaturas operativas de la GPU en niveles óptimos.
- **Espacio y Ventilación:** Un chasis con buen flujo de aire y suficiente espacio para alojar una tarjeta gráfica de tamaño completo.

Fuentes de Ayuda y Formación

Para facilitar la adopción de la AMD RX 7900 XTX, el equipo cuenta con diversas fuentes de ayuda y formación:

- **Documentación Técnica de AMD:** AMD proporciona documentación detallada, incluyendo manuales y guías de instalación, disponibles en su sitio web oficial.
- **Soporte Técnico:** Los usuarios pueden acceder al soporte técnico de AMD a través de su página web, con opciones de asistencia por chat, correo electrónico, y foros de usuarios.
- **Comunidades y Foros:** Foros como **Reddit** y **Linus Tech Tips** ofrecen una comunidad activa de usuarios y expertos que pueden ayudar con consejos y resolución de problemas.
- **Cursos y Tutoriales en Línea:** Plataformas como **Udemy**, **Coursera**, y **YouTube** ofrecen cursos y tutoriales sobre cómo optimizar y utilizar tarjetas gráficas AMD para tareas específicas, como VFX y gaming.
- **Certificaciones de Socios:** Algunos socios de AMD, como MSI y ASUS, ofrecen programas de certificación y formación específicos para sus productos, que incluyen el uso avanzado de tarjetas gráficas.

Cierre y Conclusiones

Este radar tecnológico ofrece una instantánea del estado actual de varias tecnologías emergentes que están impactando de manera significativa el sector audiovisual, con un enfoque particular en las soluciones impulsadas por la inteligencia artificial. A través de este análisis, se busca proporcionar a las empresas y profesionales del sector una guía confiable que les permita tomar decisiones informadas y apostar por soluciones tecnológicas que verdaderamente aporten valor y fortalezcan su posición en un mercado cada vez más competitivo.

Es importante destacar que el entorno tecnológico en el que nos movemos es extremadamente dinámico y está en constante evolución, con avances que ocurren a diario. Este documento no pretende ser una representación exhaustiva de todos ellos, sino ofrecer un conjunto de indicios y tendencias clave que permitan anticipar hacia dónde se dirige la tecnología en el ámbito audiovisual, ya no solo de cara a orientar a las empresas del sector en estos nuevos avances y capacidades tecnológicas, sino también como herramienta de ayuda para las administraciones, ayudándoles frente a los retos que la tecnología les presenta en su papel regulador.

Esperamos que estas pistas sean de utilidad para quienes buscan mantenerse a la vanguardia y aprovechar las oportunidades que brinda el progreso tecnológico. Debido a este carácter dinámico y avance exponencial de los desarrollos tecnológicos, recomendamos dar continuidad a este informe en el futuro. Es crucial seguir observando los movimientos y desarrollos de las herramientas analizadas en este radar, así como incorporar nuevas soluciones potenciales que puedan beneficiar al sector. Mantener esta evaluación actualizada permitirá a las empresas y a los profesionales adaptarse con agilidad a los cambios y seguir liderando la innovación en un entorno tan competitivo y transformador.

Finalmente, expresamos nuestro más sincero agradecimiento a los contribuyentes de este documento, profesionales cuya experiencia y conocimientos han sido fundamentales para la realización de este documento. Su colaboración ha sido imprescindible en la creación de este radar tecnológico que esperamos sirva como un recurso valioso para el sector de la animación, vfx y videojuegos.

Referencias

1. Motion Tracking Asistido por IA

[1] Qin, Z. and Zho, S. (2022) Object detection and tracking using Deep Learning and artificial intelligence for video surveillance applications , Research Gate. Available at: https://www.proquest.com/openview/26adfa3bc7e3508c43ff67dce050085f/1?pq-origsite=gscholar&cbl=5444811

2. Captura de Entornos con Radiance Fields

[1] Mildenhall, B., Srinivasan, P. and Tancik, M. (2020) NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, arxiv. Available at: https://arxiv.org/pdf/2003.08934

[2] Lacroute, P. and Levoy, M. (1994) *Fast volume rendering using a shear-warp factorization of ...*, *Computer Graphics Proceedings, Annual Conference Series*. Available at: http://www-graphics.stanford.edu/papers/shear/lacroute-shearwarp-sig94.pdf

[3] Kerbl, B. *et al.* (2023) *3D Gaussian Splatting for Real-Time Radiance Field Rendering*, *3D Gaussian Splatting*. Available at: https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting

[4] *NeRFstudio* (no date) *NeRFstudio*. Available at: https://docs.nerf.studio

[5] *Volinga Creator*. Available at: https://volinga.ai/

[6]*Unreal Engine*. Available at: https://www.unrealengine.com/en-US

[7] *Pixotope*. *Virtual Production Software & Solutions*. Available at: https://www.pixotope.com/

[8] *Disguise*. Available at: https://www.disguise.one/en

[9] Papantonakis, P. *et al. Reducing the Memory Footprint of 3D Gaussian Splatting*, *Symposium on Interactive 3D Graphics and Games (I3D) 2024*. Available at: https://repo-sam.inria.fr/fungraph/reduced_3dgs/

[10] Liu, Y. *et al.* (2024) *CityGaussian: Real-time High-quality Large-Scale Scene Rendering with Gaussians*. Available at: https://dekuliutesla.github.io/citygs/

[11] Li, M. *et al.* (2024) *GaussianBody: Clothed Human Reconstruction via 3d Gaussian Splatting*. Available at: https://arxiv.org/pdf/2401.09720.pdf

[12] Yang, Z. *et al.* (2024) *Real-time photorealistic dynamic scene representation and rendering with 4d Gaussian Splatting*, *4D Gaussian Splatting*. Available at: https://fudan-zvg.github.io/4d-gaussian-splatting/

[13] Dai, P. *et al.* (2024) *High-Quality Surface Reconstruction using Gaussian Surfels*, *Gaussian Surfels*. Available at: https://www.unrealengine.com/en-US

[14] *Gaussian Splatting* (no date) *Github*. Available at: https://github.com/graphdeco-inria/gaussian-splatting

[15] Yurkova, K. (2024) *A comprehensive overview of Gaussian Splatting*, *Medium*. Available at: https://towardsdatascience.com/a-comprehensive-overview-of-gaussian-splatting-e7d570081362

[16] Zhang1, D. *et al.* (2024) *Gaussian in the Wild: 3D Gaussian Splatting for Unconstrained Image Collections*, *GS-W*. Available at: https://eastbeanzhang.github.io/GS-W/

[17] Červený, J. (2023) *gsplat — 3D Gaussian Splatting WebGL viewer*, *Gsplat*. Available at: https://gsplat.tech/

3. Limpieza de planos Asistido por IA Generativa

[1] *IOPaint free toool for inpainting*. Available at: https://www.iopaint.com

[2] *LaMa: Resolution-robust Large Mask Inpainting with Fourier Convolutions*. Available at: https://github.com/advimman/lama

[3] *MAT: Mask-Aware Transformer for Large Hole Image Inpainting (CVPR 2022 Best Paper Finalist, Oral)*. Available at: https://github.com/fenglinglwb/MAT

[4] *MI-GAN: A Simple Baseline for Image Inpainting on Mobile Devices*. Available at: https://github.com/Picsart-AI-Research/MI-GAN

[5] *Stable-Diffusion-Inpainting*. Available at: https://huggingface.co/runwayml/stable-diffusion-inpainting

[6] *BrushNet*. Available at: https://github.com/TencentARC/BrushNet

[7] *ECCV 2024 | PowerPaint: A Versatile Image Inpainting Model*. Available at: https://github.com/open-mmlab/PowerPaint

[8] *AUTOMATIC1111/stable-diffusion-webui tool and documentation*. Available at: https://github.com/AUTOMATIC1111/stable-diffusion-webui

[9] *Fooocus tool and documentation*. Available at: https://github.com/lllyasviel/Fooocus

[10] *Hugging Face AI Community portal*. Available at: https://huggingface.co

[11] *Civitai AI Community portal*. Available at: https://civitai.com

4. RunwayML

[1] *Terms of Use Agreement - RunwayML* (2024) *RunwayML*. Available at: https://runwayml.com/terms-of-use

[2] *Privacy Policy - RunwayML* (2023) *RunwayML*. Available at: https://runwayml.com/privacy-policy

[3] *Pricing - RunwayML*. Available at: https://runwayml.com/pricing

[4] *Remove background – Runway*. Available at: https://help.runwayml.com/hc/en-us/articles/19112532638995-Remove-Background

5. SequoIA

[1] (2024) *Ilux Visual Technologies*. Available at: https://www.ilux.es/

6. LUMA AI Interactive Scenes

[1] *Luma AI. App Store* (2024). Available at: https://apps.apple.com/us/app/luma-ai/id1615849914

[2] *Luma AI. Google Play* (2024). Available at: https://play.google.com/store/apps/details?id=ai.lumalabs.polar&referrer=utm_source%3Dwebsite&pli=1

[3] *Luma Unreal Engine Plugin (0.41)*. Available at: https://lumaai.notion.site/Luma-Unreal-Engine-Plugin-0-41-8005919d93444c008982346185e933a1

[4] *Luma AI Capturing Best Practices.* Available at: https://docs.lumalabs.ai/MCrGAEukR4orR9

[5] *Luma WebGL Library.* Available at: https://lumalabs.ai/luma-web-library

[6] *Luma Labs AI. Dream Machine.* Available at: https://lumalabs.ai/dream-machine

[7] *Luma AI - Terms of Service. Luma AI, INC. (2024).* Available at: https://lumalabs.ai/legal/tos

[8] *Luma AI - Privacy Policy. Luma AI, INC. (2024).* Available at: https://lumalabs.ai/legal/privacy

7. Polycam

[1] *Polycam - Terms of Services (2023)*. Available at: https://poly.cam/terms_and_conditions.pdf

[2] *Polycam - Pricing*. Available at: https://poly.cam/pricing

8. Volinga

[1] *Volinga Suite - User Manual.* Available at: https://volinga.ai/docs/index.html

9. NeRFstudio

[1] *NeRFstudio. Apache License Version 2.0 January 2004. Terms and conditions for use, reproduction, and distribution.* Available at: https://github.com/*NeRF*studio-project/*NeRF*studio?tab=Apache-2.0-1-ov-file#readme

[2] *NeRFstudio Help Page.* Available at: https://docs.nerf.studio/

10. Nuke Copycat

[1] *Privacy notice. Foundry*. Available at: https://www.foundry.com/privacy-notice

[2] *Trial and buy*. *Foundry*. Available at: https://www.foundry.com/trial-and-buy

[3] *Nuke family Tech Specs. System Requirements*. *Foundry*. Available at: https://www.foundry.com/products/nuke-family/requirements

[4] *Nuke Learn. CopyCat*. *Foundry*. Available at: https://learn.foundry.com/nuke/content/reference_guide/air_nodes/copycat.html

[5] *CopyCat, Inference & Machine Learning in NukePosted* by Mike Seymour (April 14, 2021). Available at: https://www.fxguide.com/fxfeatured/copycat-inference-machine-learning-in-nuke

11. Comfy UI

[1] *Comfy Org. Main Page (2024).* Available at: https://www.comfy.org/

[2] *Comfy UI. Features.* Available at: https://github.com/comfyanonymous/ComfyUI?tab=readme-ov-file#features

[3] *Black Forest Labs (2024).* Available at: https://blackforestlabs.ai/announcing-black-forest-labs/

[4] *August 2024: Flux Support, New Frontend, For Loops, and more! (2024).* Alex Goodwin. Available at: https://blog.comfy.org/august-2024-flux-support-new-frontend-for-loops-and-more/

[5] *GNU General Public License. Version 3 (2007).* Available at: https://github.com/comfyanonymous/ComfyUI?tab=GPL-3.0-1-ov-file#readme

[6] *Get Started Comfy UI Introduction.* Available at: https://docs.comfy.org/get_started/introduction

12. Segment Anything

[1] Kirillov, A. et al. (2023) *Segment Anything.* Available at: https://arxiv.org/pdf/2304.02643

[2] *Segment Anything. Github Project.* Available at: https://github.com/facebookresearch/segment-anything

[3] *Apache License. Version 2.0, January 2024.* Available at: https://github.com/facebookresearch/segment-anything/blob/main/LICENSE

[4] *Segment Anything. Github Issues.* Available at: https://github.com/facebookresearch/segment-anything/issues

[5] *Segment Anything. Interactive Notebooks.* Available at: https://github.com/facebookresearch/segment-anything/tree/main/notebooks

13. YOLOv8

[1] Joseph R. *et al.* (2015) *You Only Look Once: Unified, Real-Time Object Detection*. Available at: https://arxiv.org/pdf/1506.02640

[2] *Ultralytics YOLO Docs.* Available at: https://docs.ultralytics.com/es

[3] *Explore Ultralytics YOLOv8.* Available at: https://yolov8.com/

[4] *GNU Affero General Public License. Version 3, 19 November 2007.* Available at: https://github.com/ultralytics/ultralytics/blob/main/LICENSE

[5] *YOLOv8 Enterprise License. Ultralytics.* Available at: https://docs.ultralytics.com/#yolo-a-brief-history:~:text=Enterprise%20License%3A%20Designed,Ultralytics%20Licensing

[6] *YOLOv8 Setting Modification. Ultralytics.* Available at: https://docs.ultralytics.com/help/privacy/#inspecting-settings

[7] *YOLOv8 Github Code Repository. Ultralytics.* Available at: https://github.com/ultralytics/ultralytics

14. OpenCV. Optical Flow. RLOF

[1] *OpenCV.* Available at: https://opencv.org/

[2] *Apache License, Version 2.0. January 2004. Apache Software Foundation.* Available at: https://www.apache.org/licenses/LICENSE-2.0

[3] *OpenCV Github License.* Available at: https://github.com/opencv/opencv/blob/4.x/LICENSE

[4] *OpenCV. OpticalFlow Algorithms. RLOF Documentation.* Available at: https://docs.opencv.org/4.x/d4/d91/classcv_1_1optflow_1_1RLOFOpticalFlowParameter.html

[5] *OpenCV Q&A Development Forum.* Available at: https://answers.opencv.org/questions/

[6] *OpenCV Algorithm, RLOF Github Repository.* Available at: h

https://github.com/tsenst/RLOFLib

15. Stable Diffusion

[1] Robin Rombach and Patrick Esser and contributors. *Stable Diffusion License, August 2022.* Available at: https://huggingface.co/spaces/CompVis/stable-diffusion-license

[2] *Realistic Vision Model V6.0 "New Vision".* Available at: https://huggingface.co/SG161222/Realistic_Vision_V6.0_B1_noVAE

[3] *Juggernaut XL V9 + RunDiffusion Photo V2 Model Official.* Available at: https://huggingface.co/RunDiffusion/Juggernaut-XL-v9

[4] Image Pipeline. *Midjourney-Mimic Model V1.2.* Available at: https://huggingface.co/imagepipeline/Midjourney-Mimic-v1.2

[5] *Controlnet Union sdx1. ControlNet++: All-in-one ControlNet for image generation and editing.* Available at: https://huggingface.co/xinsir/controlnet-union-sdxl-1.0

[6] Hu Ye. *et al.* (2023) *IP-Adapter. IP-Adapter Model Card.* Available at: https://huggingface.co/h94/IP-Adapter

[7] *Stable Diffusion XL V1.0 Inpainting API Inference.* Available at: https://huggingface.co/stablediffusionapi/stable-diffusion-xl-1.0-inpainting-0.1

[8] Dustin Podell. *et al.* (2023) *SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis*. *Stability AI, Applied Research.* Available at: https://arxiv.org/pdf/2307.01952

[9] *Tom´s Hardware Stable Diffusion Benchmarks: 45 Nvidia, AMD and Intel GPUs Compared (2023).* Available at: https://www.tomshardware.com/pc-components/gpus/stable-diffusion-benchmarks#section-stable-diffusion-768x768-performance

[10] *SDXL AWS Documentation. Stability AI*. Available at: https://stability.ai/sdxl-aws-documentation

[11] Privacy Policy. *Stability AI (2024).* Available at: https://stability.ai/privacy-policy

[12] *Stable Diffusion XL 1.0 Base Model Card, Official Documentation.* Available at: https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0

[13] *Stable Diffusion Support, Stability AI API Support.* Available at: https://stabilityplatform.freshdesk.com/support/home

16. Nvidia RTX 4090

[1] Tim Dettmers.*Which GPUs to Get for Deep Learning: My Experience and Advice for Using GPUs in Deep Learning (2023). Tensor cores.* Available at: https://timdettmers.com/2023/01/30/which-gpu-for-deep-learning/#Matrix_multiplication_with_Tensor_Cores

[2] *Tom´s Hardware Stable Diffusion Benchmarks: 45 Nvidia, AMD and Intel GPUs Compared (2023).* Available at: https://www.tomshardware.com/pc-components/gpus/stable-diffusion-benchmarks#section-stable-diffusion-768x768-performance

[3] Jarred Walton. *Best Graphics Cards for Gaming in 2024. Nvidia GeForce RTX 4090.* Available at: https://www.tomshardware.com/reviews/best-

gpus,4380.html#:~:text=If%20you%20don't%20already,than%20the%20RTX%204080%20Super

[4] NVIDIA. *Condiciones Generales de venta (CVG).* Available at: https://media.ldlc.com/pdf/nvidia/es/es/cgv.pdf

[5] CUDA. *End user License Agreement.* Available at: https://docs.nvidia.com/cuda/eula/index.html

[6] NVIDIA. *Política de Protección de datos de NVIDIA, 21 mayo 2024.* Available at: https://www.nvidia.com/es-es/about-nvidia/privacy-policy/

[7] NVIDIA. *GeForce RTX 4090.* Available at: https://www.nvidia.com/es-es/geforce/graphics-cards/40-series/rtx-4090/

[8] WhatPSU. *Suggested PSU for Intel Xeon Silver 4516Y+ and NVIDIA GeForce RTX 4090 D*. Available at: https://www.whatpsu.com/psu/cpu/Intel-Xeon-Silver-4516Y/gpu/NVIDIA-GeForce-RTX-4090

17. Nvidia RTX A6000

[1] NVIDIA. *Condiciones Generales de venta (CVG).* Available at: https://media.ldlc.com/pdf/nvidia/es/es/cgv.pdf

[2] NVIDIA. *Política de Protección de datos de NVIDIA, 21 mayo 2024.* Available at: https://www.nvidia.com/es-es/about-nvidia/privacy-policy/

[6] NVIDIA. *NVIDIA RTX A6000. Powering the world´s highest-performing workstation. Datasheet.* Avaliable at: https://www.nvidia.com/content/dam/en-zz/Solutions/design-visualization/quadro-product-literature/proviz-print-nvidia-rtx-a6000-datasheet-us-nvidia-1454980-r9-web%20(1).pdf

[7] NVIDIA. *NVIDIA RTX 6000 Ada Generation.Performance for endless possibilities. Datasheet.* Avaliable at: https://www.nvidia.com/content/dam/en-zz/Solutions/design-visualization/rtx-6000/proviz-print-rtx6000-datasheet-web-2504660.pdf

[8] WhatPSU. *Suggested PSU for Intel Xeon Silver 4516Y+ and NVIDIA GeForce RTX 4090 D*. Available at: https://www.whatpsu.com/psu/cpu/Intel-Xeon-Silver-4516Y/gpu/NVIDIA-GeForce-RTX-4090

18. AMD RX 7900 XTX

[1] Aaryaman Vasishta. *How to accelerate AI applications on RDNA 3 using WMMA. AMD GPUOpen (2023).* Available at: https://gpuopen.com/learn/wmma_on_rdna3/

[2] TechPowerUp. *AMD Radeon RX 7900 XTX Specs.* Available at: https://www.techpowerup.com/gpu-specs/radeon-rx-7900-xtx.c3941

[3] Jarred Walton. *AMD RDNA 3 GPU Architecture Deep Dive: The Ryzen Moment for GPUs. Tom´s Hardware (2023).* Available at: https://www.tomshardware.com/news/amd-rdna-3-gpu-architecture-deep-dive-the-ryzen-moment-for-gpus#section-amd-rdna-3-2nd-generation-ray-tracing

[4] Thaddée Tyl. *Recomputing ML GPU Performance: AMD vs. NVIDIA. Espadrine Blog (2023).* Available at: https://espadrine.github.io/blog/posts/recomputing-gpu-performance.html

[5] AMD. *Acuerdos de licencia de usuario final de AMD Software.* Available at: https://www.amd.com/es/legal/eula.html

[6] AMD. *Advanced Micro Devices: Privacy (2024).* Available at: https://www.amd.com/es/legal/privacy.html